J's Lab Notebook

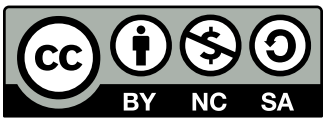
Graduate School Edition

Jeremiah Faith
Bioinformatics Program, Boston University
44 Cummington St.
Boston MA, 02215
email: faith@bu.edu

this is an *Open Science Notebook*

Notebook compiled: *April 15, 2008*

Open Notebook Science License



For more information about Open Notebook Science see the following blog articles:

- Jean-Claude Bradley coined the term *Open Notebook Science*
- Bill Hooker at 3 quarks daily has a nice series of articles about open science

About the author: I'm *J.* I'm a graduate student in Tim Gardner's lab at Boston University. For more about me see J's Page.

Notes on organization, comprehension, and raw data availability
Organization: This lab notebook is organized into chapters, which roughly correspond to the different experimental projects I am working on or have worked on in the past.

Comprehension: Prior to making this an open notebook, it was only important that myself and a few other people understand the contents of the notebook (when I want to rest of the scientific world to understand, I polish and organize the stuff into a paper). Since going open, I developed a system of minimal rules to follow so that other researchers have a better chance of understanding the big-picture of each chapter. The newer chapters abide by these rules, most of the older chapters don't yet. I'll try to update them as I go along, or at least I'll reference the paper where the work was published.

Nonetheless, the work in this document will never be as easy to understand or as flowing as a publication in a scientific journal, particularly because it has almost all of the failed experiments that happen along the way to finishing a project. You also may find many speling errors, not good grammar, and a few bad words when experiments are being a pain-in-my-ass.

Raw data availability: I switched to a latex/electronic lab notebook in September 2005. Along the way, I learned different tricks to make this thing more useful to myself. It wasn't until a year or more later that I developed a good system to link to my raw data. If you want any of the raw data that doesn't have a hyperlink to it, send me an email, I can probably dig it up from my computer.

Notes on referencing this work If for some reason you need to refer to this document in your own work *do not* refer to the page number or the section number, because they change as I add things to the notebook. I try not to change the chapter and section headings, so it's best if you refer to those. Please also include a link to the page where I keep this open science notebook.

Notes on my referencing of other work There are very few references in this document. I do give credit to the appropiate folks when chapters from this notebook are published. Since the notebook is open, I'll try to add key references as I start new chapters in the future.

Notes on completeness I do not release the chapters (currently only one) that involve work in collaboration with other labs, who may not want their research posted on the web prior to more traditional routes of publication. I also do a large amount of bioinformatics stuff and coding. In general, I don't include software in my lab notebook (for this I maintain a cvs and a wiki for tracking stuff). I currently have no plans to post my software as part of this notebook, because much of it is open source and/or published already.

Notes on authorship The large majority of the work in this notebook was done by me (Jeremiah Faith). However, I have also included "highlights" from some of the very productive rotation students that have worked with me. Details will be found in their own lab notebooks, but I'll put the main conclusions and figures in this notebook, in case they disappear to other labs after their rotations and so that it is clear where some of the conclusions and steps in this notebook derive from. To make clear work that was done by someone else I'll surround each section by any author (besides myself with) like this:

`BEGIN WORK BY FIRST_NAME LAST_NAME`
Date

work goes here
`END WORK BY FIRST_NAME LAST_NAME`

Because of the latex formatting engine some of the figures may fall outside that section, but I try reference the authors name in the figure legend.

Names of addition authors Currently the contributors of data for this manual (besides myself) are Kevin Litcofsky and Esther Rheinbay.

Contents

1 Chromatin Immunoprecipitation **24**

1.1 Chip Target Primer testing . . . 25

1.1.1 Low-throughput testing . . . 26

1.1.2 Improving gel of target genes . . . 27

1.1.3 Higher-throughput primer organization . . . 28

1.1.4 Higher-throughput primer check . . . 28

1.1.5 Higher-throughput primer check: trying for better gels . . . 30

1.2 Testing the shearing range . . . 31

1.2.1 First shear test . . . 32

1.3 Checking the cloned TOPO constructs . . . 36

1.3.1 DNA Sequencing . . . 37

1.3.2 Do any strains not have errors? . . . 43

1.3.3 Recloning TOPO constructs . . . 55

1.4 Cloning TOPO constructs into MG1655 . . . 88

1.5 ChIP optimizing parameters . . . 88

1.5.1 Round1 with lrp:A only . . . 88

1.5.2 Round1 with lrp:A2 . . . 93

1.5.3 Round1 with lrp:B1 . . . 93

1.6 lrp ChIP in Minimal and Rich media in MG1655 . . . 94

1.6.1 growing lysing shearing . . . 94

1.6.2 Misc. . . . 95

1.6.3 Bead washing . . . 96

1.6.4 Ideas for higher-throughput and automation . . . 96

1.6.5 qPCR . . . 97

1.7 ChIPrndC: pdhR, fecI, lexA, ydaK . . . 101

1.7.1 primer design . . . 103

1.7.2 qPCR . . . 103

1.8 Sharing the Transcription Factor TOPO cds cloning primers . . . 106

2 Towards a faster, more reliable ChIP protocol **108**

2.1 ChIP improvement round 1: shortening the time . . . 109

2.2 First round ChIP optimizations . . . 110

2.2.1 Phenol -vs- ChargeSwitch -vs- Qiagen PCR purification for cleanup of sheared, decrosslinked DNA . . . 110

2.2.2 first round ChIP optimizations: detailing the plan . . . 113

2.2.3 first round ChIP optimizations: growing, shearing, lysing, sonicating . . . 113

2.2.4 first round ChIP optimizations: antibody, beads, washing, crosslink reversal . 114

2.2.5 first round ChIP optimizations: purification of DNA products . . . 115

2.2.6 first round ChIP optimizations: preparing and testing the ChIP primers . . . 116

2.2.7 first round ChIP optimizations: dynamoHS qPCR . . . 117

2.2.8 first round ChIP optimizations: ABI master mix qPCR . . . 118

2.2.9 Summary of first round results . . . 119

2.3 scratchNotes . . . 120

2.4 Second round ChIP optimizations . . . 120

2.4.1 second round ChIP optimizations: detailing the plan . . . 121

2.4.2 second round ChIP optimizations: checking the sonication . . . 121

2.4.3 second round ChIP optimizations: factors . . . 122

2.4.4 second round ChIP optimizations: growing, shearing, lysing, sonicating . . . 123

2.4.5 second round ChIP optimizations: antibody, beads, washing, crosslink reversal 124

2.4.6 second round ChIP optimizations: purification of DNA products . . . 124

2.4.7 second round ChIP optimizations: dynamoHS qPCR . . . 124

2.4.8 Summary of second round results . . . 125

2.5 Response surface 1: beads vs antibody . . . 127

2.5.1 preparing the sheared chromatin . . . 127

2.5.2 testing ability of Qiagen PCR prep to work with minute quantities of DNA . 128

2.5.3 optimization of bead and antibody concentrations . . . 129

2.6 third round ChIP optimizations . . . 132

2.6.1 third round ChIP optimizations: checking the sonication . . . 134

2.7 Preparations for future factors . . . 136

2.7.1 Cloning lrpB into a low-copy plasmid . . . 136

2.8 forth round ChIP optimizations: chromatin concentration . . . 137

2.8.1 the plan . . . 137

2.8.2 grow, lyse, shear . . . 138

2.8.3 immunoprecipitation . . . 138

2.8.4 elute DNA and qPCR . . . 138

2.8.5 Summary of fourth round results . . . 139
2.9 fifth round ChIP optimizations: finishing up the bead antibody surface . . . 140
2.9.1 fifth round (bead/antibody surface) experiments . . . 140
2.9.2 Summary of fifth round (bead/antibody surface) results . . . 143
2.10 Sanity check: again, does lowing DNA concentration really allow us to lower the amount of bead and antibody . . . 144
2.11 Does the volume of the precipitation matter? . . . 145
2.12 Sanity check: does the volume of the precipitation matter? . . . 147
2.13 Making the bead/antibody surface using a small reaction volume . . . 149
2.14 Plate -vs- tube . . . 151
2.15 Making the formaldehyde/shearing surface . . . 153
2.15.1 adding more low formaldehyde concentration data points . . . 156
2.15.2 Do I really need to reverse the crosslinks overnite to quantify the amount of sheared chromatin . . . 157
2.16 testing the optimized protocol on a transcription factor besides Lrp . . . 159
2.16.1 primer plates for the three transcription factors . . . 160
2.16.2 qPCR of targets of the three transcription factors . . . 161
2.17 a factorial with two TFs: removing the lrp specific optimizations . . . 162
2.17.1 strategy to regain the lost performance . . . 162
2.17.2 running the factorial experimental design . . . 163
2.17.3 new lrp and fecI primer plates . . . 164
2.17.4 lrp and fecI fractional factorial results . . . 164
2.18 testing the 2.5 day, top-performing protocol, on Lrp, FecI, and Lrp . . . 165
2.18.1 more primer plates . . . 167
2.18.2 2.5 day protocol results . . . 168
2.19 testing the 1.5 day, top-performing protocol, on Lrp, FecI, and Lrp . . . 169
2.20 combinatorial epitope tags to improve consistency? . . . 171
2.20.1 testing the four epitope tag combinations with lexA, fecI, and lrp . . . 171
2.20.2 testing the two best epitope tag combinations with the 2.5 day protocol and more replicates . . . 175
2.20.3 expanding the new vector to new TFs . . . 177
2.21 END OF COMPLETED SECTIONS . . . 179
2.22 Cool antibody factorial experiment to maybe try someday . . . 179
2.23 Cloning lrpB into a myc-tagged plasmid . . . 179
2.23.1 miniprepping and sequencing the new lrp plasmids . . . 179
2.24 NOTES . . . 180
2.25 list of annoying things I'd like to check with a final factorial . . . 180
2.25.1 other things to check . . . 181

3 Serial Analysis of Promoter Enrichment 182
3.1 Initial Steps . . . 183
3.1.1 First SAPE Protocol . . . 185
3.1.2 SAPE1: the return of SAPE1 . . . 191
3.1.3 SAPE1 returns: following the same sample all the way (including PCR) . . . 194
3.1.4 More DNA: midipreps, midipreps, midipreps . . . 196
3.1.5 Midipreps that work and are RNA free! . . . 197
3.1.6 SAPE2 a modified protocol that hopefully works (original SAPE did not) . . . 198
3.1.7 stab at an initial SAPE2 detailed protocol . . . 198
3.2 The end of SAPE . . . 200

4 chromatin immunoprecipitation based determination of transcription factor binding sites: SAPE returns 201
4.1 Will restriction enzymes digest crosslinked DNA . . . 201
4.2 Cloning ChIP DNA . . . 203
4.2.1 preparing the enriched/cut Chromatin . . . 203
4.2.2 cloning the cut chromatin . . . 204
4.3 Cloning ChIP DNA: try 2 . . . 204

5 Experimental Operon Determination 206
5.1 Initial Tests . . . 208
5.1.1 flgK, flgL operon primer tests . . . 208
5.1.2 PCR tips . . . 210
5.2 Moving to a more realistic scale . . . 210
5.2.1 Getting started . . . 211
5.2.2 Preping genomic DNA . . . 212
5.2.3 Preping genomic DNA: 2nd try . . . 213
5.2.4 Preping genomic DNA: with shearing . . . 213
5.2.5 Preparing the oligos . . . 215
5.2.6 Realistic Scale: first experiments → single-plex . . . 215
5.2.7 Multiplexing chunks1 and 2 . . . 217
5.3 Moving on to cDNA from a few conditions . . . 223
5.3.1 Initial chosen conditions . . . 223
5.3.2 Growing cells and RNAprotect . . . 223
5.3.3 RNApreps and cDNA . . . 224
5.3.4 2-plex PCR on cDNA sample 3 (norflox) . . . 225
5.3.5 2-plex PCR on cDNA all samples . . . 227

5.4 cDNA from 4 conditions with replication . . . 227
5.4.1 Growth and RNAProtect . . . 227
5.4.2 RNA preps, DNAse digestion and cDNA . . . 229
5.4.3 384-well qPCR . . . 230
5.5 gDNA free RNA . . . 232

6 Gene and operon boundary determination by barcoded paired-end-tags and highly parallel sequencing 237
6.1 Planning and Goals . . . 238
6.1.1 Progress Reports on above enumeration . . . 239
6.2 Cloning double-stranded cDNA starting from total RNA . . . 240
6.2.1 First steps . . . 240
6.3 Cloning double-stranded cDNA from total RNA, using adaptors . . . 248
6.3.1 RNA to cDNA . . . 248
6.3.2 Preparing cDNA and vector for cloning . . . 251
6.3.3 cloning adaptored cDNA . . . 253
6.3.4 size selecting cDNA for cloning . . . 255
6.4 Cloning double-stranded cDNA from *mRNA*, using adaptors . . . 258
6.4.1 RNA to cDNA . . . 259
6.4.2 Preparing cDNA and vector for cloning . . . 261
6.5 Cloning double-stranded cDNA from *mRNA*, using adaptors and more RNA . . . 264
6.6 Circularization test with sample 3 . . . 267
6.6.1 Circularization adaptors and strategy . . . 267
6.6.2 Circularization first attempt . . . 268
6.7 Ligating the final adaptor ends for polony sequencing . . . 271
6.7.1 Ligating for cloning using circularized/digested sample 3 . . . 271
6.7.2 Ligating ends using a PCR product . . . 279
6.7.3 Ligating ends using an RCA product *again* . . . 284
6.8 Amplifying the adaptored tags . . . 292
6.8.1 PCR Primers . . . 292
6.9 Cloning the adaptored tags . . . 294
6.9.1 cloning the sticky PCR product into pUC19 . . . 294
6.9.2 checking the clones . . . 295
6.10 Cloning the unadaptored tags . . . 296
6.10.1 blunt coloning unadaptored tags try 1 . . . 296
6.10.2 blunt coloning unadaptored tags try 2 . . . 297

6.10.3 blunt coloning unadaptored tags try 3, going old-school . . . 298
6.11 New circularization adaptors . . . 301
6.12 Concatenating tags for Sanger and Pyro sequencing . . . 302
6.12.1 The adaptor sequence for concatenation . . . 302
6.12.2 Preparing cDNA for the concatenation approach . . . 303
6.12.3 Ligating tags into circles for RCA amplification . . . 307
6.12.4 amplification of tags with LM-PCR . . . 319
6.13 Further removal of rRNA . . . 320
6.13.1 Further removal of rRNA: strategy . . . 320
6.13.2 Further removal of rRNA: first try . . . 320
6.13.3 further removal of rRNA: 2nd try focus on ESTs . . . 325
6.14 How many samples must be pooled for one 454 run . . . 330
6.14.1 Amplifying the PET DNA . . . 331

7 Removal of rRNA from prokaryotic total RNA samples 333
7.1 developing a gentle, quick-lyse procedure that produces undegraded total RNA . . . 333
7.1.1 quick-and-simple readyLyse preparation of total RNA try 1 . . . 333
7.1.2 quick-and-simple readyLyse preparation of total RNA try 2 . . . 334
7.2 problems with assaying rRNA removal . . . 335
7.3 rRNA removal via RNAseH . . . 335
7.3.1 RNAseH tests . . . 336
7.3.2 can I get a decent RNA ladder gel . . . 336
7.3.3 testing RNAse H on total RNA . . . 337
7.3.4 testing RNAse H on total RNA with DNA oligos complementary to the rRNA 339
7.3.5 testing RNAse H on total RNA with DNA oligos designed by primer3 . . . 341
7.3.6 optimizing RNAse H and DNA oligo concentration . . . 343
7.3.7 optimizing RNAse H and DNA oligo concentration using reverse complement oligos . . . 344
7.3.8 testing RNAse H on total RNA with a PLATE DNA oligos complementary to the rRNA . . . 345
7.3.9 running low on total RNA . . . 348
7.3.10 testing RNAse H on total RNA with a PLATE DNA oligos complementary to the rRNA and MEGAClear . . . 348
7.3.11 cloning and sequencing to test the RNAse H based rRNA removal methods . 350
7.3.12 specificity of RNAse H oligos . . . 357
7.3.13 initial testing of the 16S mutation oligos . . . 358

7.3.14 cloning and sequencing to test additions RNAse H and oligo pulldown combinations . . . 362
7.4 rRNA removal via oligo pulldown . . . 365
7.4.1 biotinylating the oligo mix . . . 365
7.5 rRNA removal via ultracentrifugation . . . 366
7.6 rRNA removal via column chromatography . . . 366

8 USER cloning **367**
8.1 Motivation . . . 368
8.2 Plan . . . 368
8.3 Cloning araB . . . 369
8.3.1 PCR with uradine primers, Linearization of pNEB193 . . . 370
8.4 Trying with mCherry and lrp . . . 372
8.4.1 mCherry and lrp USER primer design . . . 372
8.4.2 USER PCR of mCherry and lrp . . . 373
8.4.3 USER cloning of mCherry and lrp . . . 374

9 Dabbling in synthetic biology **377**
9.1 testing pIKE 107 . . . 378
9.2 making a pyruvate / lactose regulated switch . . . 378
9.2.1 Cloning in ndh and pdhR promoters . . . 378
9.2.2 Cloning the pdhR gene into the pdhR-prom and ndh-prom vectors . . . 381
9.2.3 Ilaria's testing of the individual promoters in the U. Alon plasmid . . . 383
9.2.4 Testing the pdhR version of the toggle . . . 385

10 Miscellaneous Experiments **389**
10.1 New color proteins from the Tsien lab . . . 389
10.1.1 Filter Selection . . . 390
10.1.2 Archiving the colors . . . 390
10.1.3 Putting the colors in a strain where they'll express . . . 390
10.2 Sephacryl column separation of DNA . . . 391
10.2.1 Testing the old protocol . . . 393
10.2.2 Improving the sephacryl protocol . . . 394
10.2.3 Improving the sephacryl protocol, part 2 . . . 396
10.3 Comparison of short DNA fragment removal with Microcon 30, 50, or Qiagen PCR purification kit . . . 397
10.3.1 preparing the 80mer and 120mer PCR products . . . 397

10.3.2 preparing the test DNA . . . 397
10.4 Comparison of short DNA fragment removal with Qiagen PCR purification kit and ChargeSwitch with 1x, 2x, 3x, 4x N5 buffer . . . 398
10.5 Comparison of Qiagen Column Gel cleanup and QiaexII gel cleanup . . . 400
10.6 Comparison of short DNA fragment removal with PCR Purelink kit from Invitrogen 401
10.7 diffusion of DNA when loading agarose gels . . . 402
10.8 Comparison of DNA loss with Qiagen DNA purification vs Qiagen MinElute DNA purification . . . 403

A Techniques **405**
A.1 Nucleic Acid Determination by Spectrophotometry . . . 405
A.2 Pipetting . . . 405
A.2.1 Multichannel pipetting . . . 405

B Recipes **407**
B.1 Antibiotic stocks . . . 407
B.1.1 Norfloxacin / Ciprofloxacin . . . 407
B.2 Common Buffers . . . 407
B.2.1 Phosphate Buffered Saline (PBS) . . . 407
B.3 Agarose gel buffers what? . . . 407
B.3.1 Tris-Acetate-EDTA (TAE) . . . 407
B.3.2 Tris-Borate-EDTA (TBE) . . . 408
B.3.3 Sodium Boric Acid (SB) . . . 408
B.4 DNA and RNA storage buffers . . . 408
B.4.1 Annealing of two oligos . . . 408
B.4.2 Resuspension of lypholized primers . . . 408
B.5 Midiprep Solutions . . . 408
B.5.1 Alkaline Lysis Solution 1 . . . 409
B.5.2 Alkaline Lysis Solution 2 . . . 409
B.5.3 Alkaline Lysis Solution 3 . . . 409
B.6 ChIP Recipes . . . 409
B.6.1 Palsson lysis buffer . . . 409
B.6.2 Palsson 2x IP buffer . . . 409
B.6.3 Dilution Buffer . . . 410
B.6.4 Low Salt Immune Complex Wash Buffer . . . 410
B.6.5 High Salt Immune Complex Wash Buffer . . . 410
B.6.6 LiCl Immune Complex Wash Buffer . . . 410

C Protocols **411**

C.1 Reverse transcription . . . 411
C.1.1 Superscript II/III . . . 411
C.1.2 How much? . . . 411
C.1.3 Heat stability . . . 411
C.2 Cloning . . . 411
C.2.1 Buying chemically competent cells . . . 411
C.2.2 Cloning vector into purchased competent cells . . . 411
C.2.3 Cloning vector into purchased competent cells . . . 412
C.2.4 Chemically competent cell preparation . . . 413
C.3 ChIP Protocols . . . 413
C.3.1 Lyse, Crosslink, Shear . . . 413
C.3.2 Sheared DNA yield / Beginning of immunoprecipation . . . 414
C.3.3 Immunoprecipitate Bead washing . . . 416
C.3.4 Final DNA cleanup . . . 416
C.3.5 qPCR to determine TF binding site enrichment . . . 417
C.4 ChIP Protocol Post 1st Round Factorial Optimization . . . 417
C.4.1 Lyse, Crosslink, Shear . . . 417
C.4.2 Sheared DNA yield, immunoprecipation, and bead washing . . . 418
C.5 Preparation of *E. coli* genomic DNA . . . 420
C.6 Preparation of *E. coli* plasmid DNA . . . 422
C.6.1 Standard Method . . . 422
C.6.2 Old School Method . . . 422
C.6.3 RNA-free Midiprep . . . 423
C.6.4 Midiprep . . . 424
C.7 Preparation of *E. coli* RNA . . . 425
C.7.1 RNAeasy preps . . . 425
C.8 Size-separation / exclusion of DNA . . . 426
C.8.1 Size-exclusion of DNA using microcon filters . . . 426
C.8.2 Size-exclusion using Qiagen spin-columns . . . 426
C.8.3 Size-separation of DNA using Sephacryl 500 . . . 426
C.9 Preparation of PET libraries . . . 427
C.9.1 Growing cells . . . 427
C.9.2 RNA extraction . . . 427
C.9.3 1st strand synthesis of cDNA . . . 428
C.9.4 2nd strand synthesis of cDNA . . . 428

C.9.5 Add adaptors to double-stranded cDNA . . . 429
C.9.6 Size-selection of cDNA . . . 430

D Equipment **431**
D.1 DNA work . . . 431
D.1.1 qPCR . . . 432
D.2 Gels . . . 432
D.2.1 MiniGels . . . 432
D.2.2 Big Gels . . . 433
D.2.3 Polyacrylamide pre-cast Gels . . . 433
D.3 Imaging . . . 434
D.4 Weighing, Measuring, pH . . . 434
D.5 Growing stuff and warming stuff . . . 434

List of Figures

1.1 Questions ChIP results should help answer. 25

1.2 80 ml, 1% agarose gel with 2.5 ul of 1% ethidium bromide run for 40 min at 100 volts. 27

1.3 70 ml, 2% agarose gel run for 40 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 40 min followed by 15 min in water alone (both on orbital shaker at 50rpm) to reduce background. 28

1.4 200 ml, 2% agarose gel with 4 ul of 1% ethidium bromide run for 45 min at 120 volts. 29

1.5 70 ml, 2% agarose gel run for 45 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 45 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background 30

1.6 70 ml, 2% agarose gel run for 38 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 45 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. 31

1.7 70 ml, 2% agarose gel run for 37 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 45 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. 32

1.8 70 ml, 2% agarose gel run for 34 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 45 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. 33

1.9 amount of sheared DNA obtained from lysate as a function of the OD of the culture it was extracted from. 35

1.10 70 ml, 1.5% agarose gel run for 48 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 32 min followed by 10 min in water alone (both on orbital shaker at 50rpm) to reduce background. 36

1.11 70 ml, 1.5% agarose gel run for 54 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 40 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. 37

1.12 70 ml, 2% agarose gel run for 38 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 40 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. 38

1.13 70 ml, 2% agarose gel with 0.75 ul of 1% ethidium bromide run for 40 min at 120 volts. 39

1.14 Potential PCR cloning mutation 40

1.15 70 ml, 1.5% agarose gel run for 40 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 45 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. 45
1.16 300 ml, 1% agarose gel with 1.5 ul of 1% ethidium bromide run for 38 min at 120 volts. 55
1.17 250 ml, 1.5% agarose gel with 1.5 ul of 1% ethidium bromide run for 45 min at 120 volts. 57
1.18 300 ml, 1.8% agarose gel with 1.5 ul of 1% ethidium bromide run for 48 min at 120 volts. 60
1.19 300 ml, 1.8% agarose gel with 1.5 ul of 1% ethidium bromide run for 48 min at 120 volts. 61
1.20 chromatagraphs for hyaC La and pdhR Ha 73
1.21 300 ml, 1.5% agarose gel with 4.5 ul of 1% ethidium bromide run for 40 min at 100 volts. 75
1.22 80 ml, 1.5% agarose gel with 0.4 ul of 1% ethidium bromide run for 40 min at 100 volts. 88
1.23 80 ml, 1.5% agarose gel with 0.4 ul of 1% ethidium bromide run for 40 min at 100 volts. 89
1.24 80 ml, 1% agarose gel with 0.5 ul of 1% ethidium bromide run for 40 min at 100 volts. 90
1.25 Abi 384-well qPCR plate 91
1.26 80 ml, 1.5% agarose gel run for 40 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 40 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. 95
1.27 98
1.28 300 ml, 1.5% agarose gel with 4.5 ul of 1% ethidium bromide run for 40 min at 100 volts. 99
1.29 making sure pdhR, fecI, and lexA were properly labeled 102
1.30 lexA, pdhR, fecI sheared DNA gel 103
1.31 300 ml, 1.5% agarose gel with 1.8 ul of 1% ethidium bromide run for 40 min at 100 volts. 104
1.32 300 ml, 1.5% agarose gel with 1.8 ul of 1% ethidium bromide run for 40 min at 100 volts. 105

2.1 The qiagen PCR cleanup kit is the only DNA cleanup method that was linear across the dilution range 112
2.2 The phenol:chloroform extraction lanes retained the tiny DNA fragments in the high volume purification, which may have led to the strange non-linear scaling of the measured concentrations. 112
2.3 1.5% gel; 16 sheared samples; the lightly sheared samples were 1x10%x30secs. the others were 4x20%x30secs. 115

2.4 2% gel of the random primers and lrp primers to be used in the factorial ChIP optimization experiment. The identity of the lanes is the same as in the table Lrp factorial primer plate table above. . . . 117
2.5 shearing range tests with different concentrations of DNA and formaldehyde, different shearing amounts, and different buffers. gel is a 1.5% agarose gel. . . . 123
2.6 1.5% gel run 45 min at 120V; 16 sheared randomized samples from factorial round 2 125
2.7 Sheared chromatin on agarose gel. This sheared DNA will be used for the response surface of beads/antibody. . . . 127
2.8 fill in . . . 129
2.9 ChIP enrichment of positive control genes relative to random genes for bead/antibody concentrations determined via steepest assent. . . . 131
2.10 new lrp, random, cysB, and lexA primers amplified from sheared DNA . . . 132
2.11 comparison of a newly synthesized set of lrp target primers relative to the performance with the old set I've been using up until now . . . 133
2.12 Six sheared samples on 1% agarose gel. . . . 135
2.13 Enrichment (the z-axis displayed via color) certainly is dependent on antibody and bead concentrations. There appears to be some saturation (too much DNA?). . . . 135
2.14 gel of lrp PCR prior to cloning into low-copy plasmid . . . 136
2.15 figure summarizing results of the 4 factorial experiment. Note that I've shown the forth dimension (DNA concentration) by putting the points right next to each other. The low DNA concentration is on the left and the high DNA concentration is on the right. This makes it appear that the low DNA sample has slightly less beads than the high DNA sample, this isn't the case. I just change the amounts slightly on purpose so the datapoints wouldn't overlap. For an alternative view see Figure 2.16 . . . 139
2.16 figure summarizing results of the 4 factorial experiment. this figure is plotted in 3d for an alternative view to Figure 2.15 for livK. . . . 139
2.17 . . . 141
2.18 Contour plots from round 5 . . . 143
2.19 Sheared chromatin for four DNA concentrations using two combinations of bead and antibody . . . 145
2.20 Enrichment when using a series of different volumes for the precipitation reaction . . 146
2.21 Six samples where sheared, crosslink reversed, and cleaned up. 5 μl of each (appx 600 ng) were run in each lane. The average shearing length is around the expected 300-400 bp range. . . . 148
2.22 Combined data from the two volume scan replicates. All points are the average of two replicates *except* the two points at 251 μl . These each (i.e. 1:50 and 2:100 bead:antibody) had a screwy outlier (>2x log units away from all other points) that I removed. . . . 149
2.23 Bead/Antibody enrichment surface for high and low volume samples. The A) high volume samples panel is taken from Figure 2.18 but with the addition of the actual datapoints used to fit the contour. . . . 151

2.24 Bead/Antibody enrichment surface for high and low volume samples. The A) high volume samples panel is taken from Figure 2.18 but with the addition of the actual datapoints used to fit the contour. This plot is exactly the same data as Figure 2.23 except that both the high volume and the low volume are plotted using the same scale for the colorbar. . . . 152
2.25 Sheared chromatin for different formaldehyde and bead concentrations . . . 154
2.26 It seems that lower formaldehyde concentrations are enriching better. . . . 155
2.27 Crosslink for all 8 samples were reversed at 65C with proteinase K overnite (A) or for 1 hr (B). . . . 157
2.28 Further confirmation that the lower formaldehyde concentrations are performing best 158
2.29 1.5% agarose gel of sheared chromatin run for 37 minutes . . . 167
2.30 Boxplot of the old protocol vs the 2.5 day protocol. . . . 168
2.31 1.5% agarose gel of sheared chromatin run for 30 minutes . . . 170
2.32 Boxplot of the old protocol vs the 2.5 day vs the 1.5 day protocol. . . . 170
2.33 Ilaria made tagged versions of lexA, fecI, and lrp using four different locations and combinations of the Xpress tag and the myc tag . . . 171
2.34 . . . 177
2.35 gel of cut lrp lowcopy and cut lrp myc . . . 180

3.1 SAPE: first (failed) attempt . . . 187
3.2 SAPE: first (failed) attempt ligation . . . 188
3.3 SAPE: first (failed) attempt ligation2 . . . 189
3.4 SAPE: first (failed) attempt found 2 . . . 190
3.5 SAPE1: correct digestion . . . 191
3.6 SAPE1: the return of SAPE1 part 1 . . . 193
3.7 SAPE1 returns: failed midiprep . . . 194
3.8 qiagen midipreps . . . 196
3.9 old school midipreps . . . 197
3.10 midipreps that work! . . . 198
3.11 SAPE2: schematic overview . . . 199

4.1 A schema for highly-parallel discovery of transcription factor binding sites. . . . 202
4.2 gel of cut pUC19 with different enzymes and different concentrations of formaldehyde used for crosslinking . . . 203

5.1 Experimental Operon Validation Strategy . . . 207
5.2 80 ml, 2% agarose gel with 0.5 ul of 1% ethidium bromide run for 40 min at 100 volts. 209
5.3 Experimental Gene Set . . . 211
5.4 genomic DNA . . . 214

5.5 sheared genomic DNA . . . 214
5.6 1-plex operon chunk1 on gel (failed) . . . 216
5.7 uneven wells on 96-well PCR plate . . . 216
5.8 1-plex working . . . 218
5.9 multiplex kinda working . . . 219
5.10 multiplex kinda working . . . 220
5.11 versadoc example . . . 221
5.12 TBE gel . . . 222
5.13 TBE gel . . . 222
5.14 Operon Condition set 1 RNA . . . 225
5.15 Sample 2 cDNA operon . . . 226
5.16 Sample 2 cDNA operon . . . 233
5.17 Sample 2 cDNA operon . . . 234
5.18 operon 4 conditions 3 replicates . . . 235
5.19 Sample 2 cDNA operon . . . 236

6.1 . . . 238
6.2 . . . 239
6.3 cDNA 2 conditions 3 replicates . . . 242
6.4 RNA and cDNA 2 conditions . . . 245
6.5 cDNA clone inserts . . . 248
6.6 subheading . . . 252
6.7 cDNA samples 3 and 4 transformation . . . 254
6.8 cDNA clone inserts samples 3 and 4 no-size select . . . 254
6.9 cDNA samples 3 . . . 256
6.10 size selection of cDNA sample 3 . . . 256
6.11 sized selected cDNA clone inserts for sample 3 . . . 258
6.12 subheading . . . 262
6.13 size selection of cDNA samples 5 and 6 . . . 263
6.14 insert check cDNA samples 5 and 6 . . . 264
6.15 size select cDNA samples 3 and 6 . . . 265
6.16 insert check cDNA sample 5 . . . 265
6.17 insert check cDNA samples 3 and 6 . . . 266
6.18 circularization schema . . . 268
6.19 circularization cut with mmeI attempt 1 . . . 269
6.20 cDNA sample 3big mmeI digested . . . 271

6.21 . . . 273
6.22 yield from mmeI size-selection gel . . . 273
6.23 a) the ligation product (No PCR) and the PCR producted from amplifying 0.5 μl of the ligation product. b) zoomed in on the No PCR lane . . . 275
6.24 primers on polyacrylamide . . . 276
6.25 Urea gel with wrong dye . . . 277
6.26 primers Urea gel . . . 277
6.27 primers Urea gel2 . . . 278
6.28 primers Urea gel3 . . . 280
6.29 PCR ligation urea . . . 281
6.30 PCR ligation urea 2 . . . 282
6.31 PCR ligation urea 3 . . . 283
6.32 PCR ligation TBE1 . . . 284
6.33 PCR ligation TBE2 . . . 285
6.34 PCR ligation TBE2 . . . 286
6.35 gel lane floated away . . . 287
6.36 RCA mmeI digest on 6% TBE . . . 288
6.37 RCA mmeI digest, adaptored on 6% TBE . . . 289
6.38 polyacrylamide adaptered PETs . . . 290
6.39 Nusieve gel of adaptered PETs . . . 291
6.40 Nusieve gel of adaptered PETs . . . 292
6.41 PCR of cut Nusieve slice . . . 293
6.42 PCR of cut Nusieve slice . . . 294
6.43 PCR from cloned PET minipreps . . . 295
6.44 RCA2b tag was cut from lanes 3 and 4 and combined into one Qiagen gel cleanup . . . 296
6.45 PCR from cloned PET minipreps . . . 297
6.46 PET ZeroBlunt try 2 . . . 298
6.47 PET blunt pUC19 . . . 300
6.48 PET contatenation schema . . . 302
6.49 mRNA growth LB and LB + norflox . . . 304
6.50 RNA samples 1 and 4 for PET . . . 305
6.51 cDNA samples 1 and 4 . . . 306
6.52 cDNA samples 1 and 4 mmeI digest . . . 307
6.53 83.2 circulize test1 . . . 311
6.54 83.1 circulize test1 . . . 313
6.55 83.1 gel exonuclease . . . 314

6.56 83.1 gel quantify . . . 315
6.57 83.1 gel exonuclease . . . 316
6.58 2% gel run for 45 minutes. Unfortunately, I don't see the expected band at 83 bp. . . . 317
6.59 83.2 gel exonuclease . . . 318
6.60 2% gel run for 45 minutes. Seems lane 1 worked and lane 2 didn't. . . . 319
6.61 1.5% gel run for 45 minutes. . . . 320
6.62 total RNA was run through the MICROBExpress kit one time and two times (2x) to remove the rRNA. this is a 1.5% TAE agarose gel . . . 321
6.63 . . . 324
6.64 . . . 325
6.65 . . . 327
6.66 . . . 329
6.67 The lanes from top to bottom are MmeI cut A.2x, bank, 4 μl 25 bp ladder (invitrogen), NEB 2-log ladder, 2 μl 25 bp ladder (invitrogen), MmeI cut A.1x. . . . 332

7.1 . . . 335
7.2 . . . 337
7.3 . . . 337
7.4 . . . 339
7.5 . . . 340
7.6 . . . 342
7.7 . . . 344
7.8 . . . 345
7.9 . . . 347
7.10 . . . 348
7.11 . . . 350
7.12 . . . 350
7.13 . . . 353
7.14 . . . 359
7.15 . . . 361
7.16 . . . 365

8.1 USER cloning Schema . . . 369
8.2 USER cloning PCR araB . . . 371
8.3 USER cloning PCR lrp and mCherry . . . 374
8.4 USER cloning PCR lrp and mCherry insert check . . . 375

9.1 network inference toggle overview . . . 378

9.2 network inference toggle . . . 379
9.3 cut vector and inserts for pIKE . . . 380
9.4 insert check for pIKE and nIKE . . . 381
9.5 cut vector and inserts for pIKE and nIKE . . . 382
9.6 cut vector and inserts for pIKE and nIKE . . . 383
9.7 toggle promoter pre-screen Alon plasmid . . . 383
9.8 GFP estimated expression vs qPCR estimated expression . . . 384
9.9 GFP estimates of ndh and pdhR promoters . . . 385
9.10 toggling the new toggle . . . 387

10.1 plate reader experiment with Tsien colors . . . 392
10.2 I washed the cells, from an overnight in the BL21(DE3)pLyseE strain with 500 uM IPTG, with Davis minimal media. Image (a) shows proteins mCherry, Citrine, and mOrange (left to right) in suspension. Image (b) shows the same tubes but with the cells pelleted to concentrate the proteins/cells to one location (left to right is Citrine, mCherry, mOrange). Image (c) shows pellets of cultures expressing mCitrine, mOrange, and mCherry in white (ambient) light. . . . 393
10.3 500 μl of Sephacryl S500 (a) was washed 3x with TE. The centrifugal force distributed the sephacryl on a hard angle inside the Spin-X column (b). . . . 394
10.4 . . . 394
10.5 850 μl sephacryl . . . 395
10.6 . . . 395
10.7 sephacryl gel 1 ml and 1.2 ml . . . 396
10.8 pLtet 80mer 120mer PCR . . . 398
10.9 . . . 399
10.10 . . . 399
10.11 . . . 402
10.12 . . . 403

D.1 PCR machine . . . 431
D.2 . . . 435
D.3 PCR machine . . . 436
D.4 The Denver 200 pH meter does the trick . . . 436
D.5 This Thermo electron shaking incubator is very quiet but only works well at temperatures 35C and above. It is also *very slow* to start and stop and has an annoyingly long delay between when you push the start/stop button and when it decides "ok I'll begin stopping now". We have two stackable ones. . . . 437

List of Tables

1.1 Primer organization 29

1.2 Oligo concentration is 4μM. Wells contain both a forward and reverse primer. lexA contains the corrected lexA reverse primer. gfp requires a different template (since it is not in genome it must be amplified from plasmid). Oligo names with -m indicate the myc version of the primer that removes the stop codon (necessary for proper production of the myc tag). bolA-m didn't make it into the final plate because I forgot to order the primer. 56

1.3 L = low conc. plate (plated 10μl of ligation reaction), H = high conc. plate (35μl of ligation), † = incorrect/no insert in digest (see Figure 1.18) 58

1.4 lrp:A1 qPCR plate setup. Yellow rows are + control template (sheared lysate DNA). Cyan rows are immunoprecipitated DNA with a specific antibody. Red are immunoprecipitated with an unrelated antibody. White are immunoprecipitated with no antibody. 91

1.5 qPCR Ct values rounded to the nearest 1000th. pCnt = positive control, sheared lysate not precipitated; sAnti = target DNA, immunoprecipitated with the antibody specific to the TF of interest; uAnti = target DNA, immunoprecipitated with an unrelated antibody (myc for an XPress tagged protein), nAnti = beads only, no antibody control. All samples except the sAnti have four replicates. One of the sAnti replicates was lost by pipetting into the wrong hole. 92

1.6 lrp:A1 qPCR plate setup. Yellow rows are + control template (sheared lysate DNA). Cyan rows are immunoprecipitated DNA with a specific antibody. Red are immunoprecipitated with an unrelated antibody. White are immunoprecipitated with no antibody. Blanks are template + master mix only, no primers. 5kUP and 1kbUP were added to see how good the shearing is. pntA is a new potential target from Boris' algorithm. gltB and stpA are new extra positive controls. 93

1.7 yellow is positive control, cyan = A (correct antibody), magenta = B (incorrect antibody), white = C (beads only) 99

Chapter 1

Chromatin Immunoprecipitation

THIS CHAPTER IS COMPLETE

Brief Update *Thu Jul 12 18:54:37 EDT 2007*: The ChIP verifications done in this chapter were published in the PLoS Biology paper: Large-Scale Mapping and Validation of Escherichia coli Transcriptional Regulation from a Compendium of Expression Profiles. Please see that paper for background, introduction, and a discussion of this work in the context of our work on network inference. For this work in particular, see the section of that paper entitled *In vivo confirmation of new regulatory interactions*. The detailed protocol that resulted from the work in this chapter is available in the appendix section C.3 on page 413 with recipes available in section B.6 on page 409.

I'm trying to verify predicted interactions using ChIP.

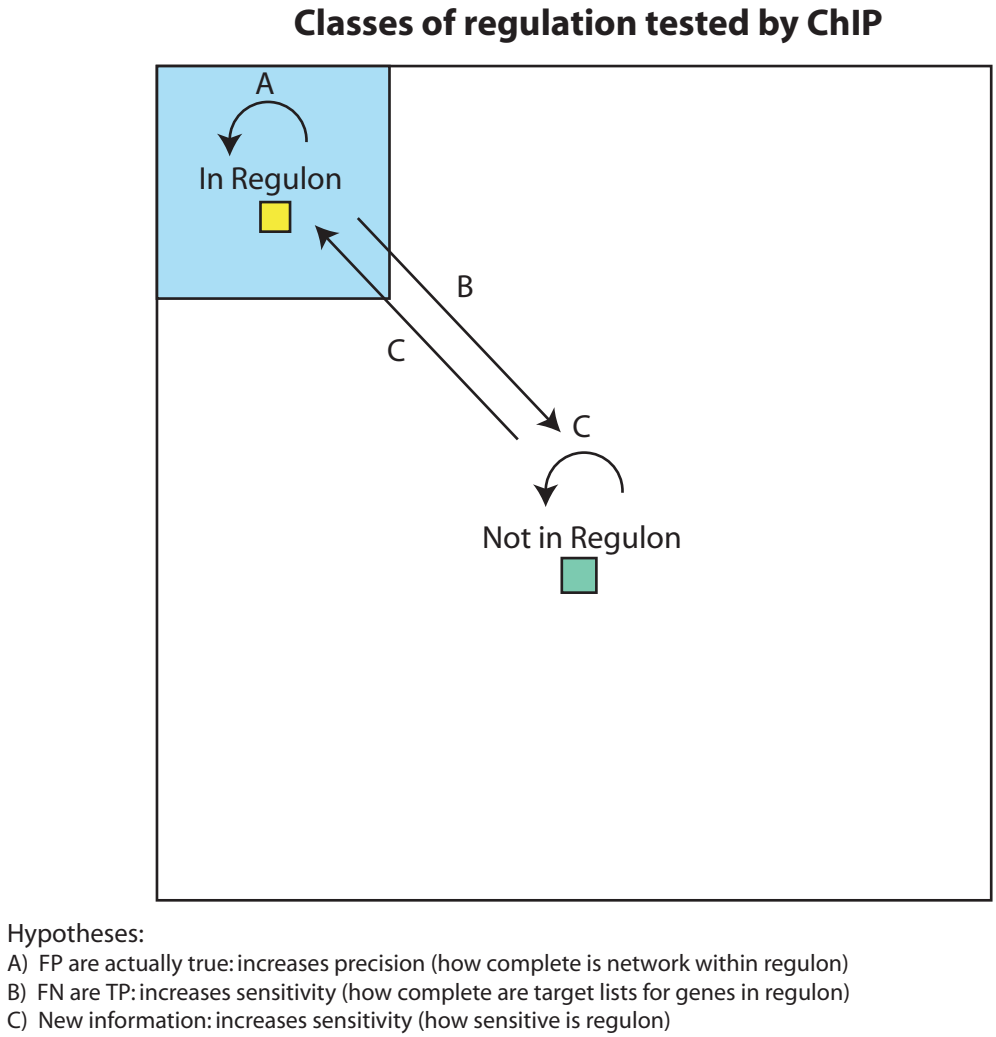


Figure 1.1: Figure is drawn to scale. Big box is number of genes in regulon (genes only). Blue box is size of regulon. Space not covered by blue box is all non-regulon genes. Green box is number of TFs outsize regulon. Yellow is number of TFs in regulon.

1.1 Chip Target Primer testing

For each predicted target of the transcription factors to be detected we designed a primer pair to amplify the intergenic region and test if the region has been enriched by immunoprecipitation by an antibody directed at the transcription factor of interest. Primers were designed using Primer3 software with the following constaints: length 24bp+/-3, melting temp 60+/-3, %GC 50+/-8. Here we are verifying that all primers amplify only one region of the correct length.

1.1.1 Low-throughput testing

Wed Sep 28 13:57:27 EDT 2005 testing 8 genes (aroG, aroL, aroP, asd, cirA, cysC, dinD) also testing two efficiency parameters:

1. can we run a 2-step rather than a 3-step reaction to speed simple primer design tests
2. can we use half of qiagen's recommended 100ul reaction volume to save reagents

all eight genes are run in a 50ul reaction (half the qiagen recommended reaction):

PCR Reaction composition	
H_2O	23.5 μl
Qiagen Master Mix	25 μl
Forward and reverse primer	1.5 μl
final primer concentration	N pM

Thermal cycler conditions		
Initial denaturation	5 min	95°C
2-Step cycling		
Denature:	30 sec	95°C
Anneal/Extend:	45 sec	60°C
Number of Cycles:	25	
Final Extention:	7 min	72°C

aroG and aroL are included as a 100ul reaction (the recommended solution) the PCR parameters are:

PCR Reaction composition	
H_2O	47 μl
Qiagen Master Mix	50 μl
Forward and reverse primer	3 μl
final primer concentration	N pM

Thermal cycler conditions		
Initial denaturation	5 min	95°C
2-Step cycling		
Denature:	30 sec	95°C
Anneal/Extend:	45 sec	60°C
Number of Cycles:	25	
Final Extention:	7 min	72°C

additionally aroG, aroL, aroP, asd were run as in a 3-step 50ul reaction:

PCR Reaction composition	
H_2O	2 μl
Qiagen Master Mix	5 μl
Forward and reverse primer	23 mM

Thermal cycler conditions		
Initial denaturation	5 min	95°C
3-Step cycling		
Denature:	30 sec	95°C
Anneal:	30 sec	60°C
Extend:	30 sec	72°C
Number of Cycles:	30	
Final Extention:	7 min	72°C

aroP and asd were run as 100ul reactions:

PCR Reaction composition	
H_2O	47 μl
Qiagen Master Mix	50 μl
Forward and reverse primer	3 mM

Thermal cycler conditions		
Initial denaturation	5 min	95°C
3-Step cycling		
Denature:	30 sec	95°C
Anneal:	30 sec	60°C
Extend:	30 sec	72°C
Number of Cycles:	25	
Final Extention:	7 min	72°C

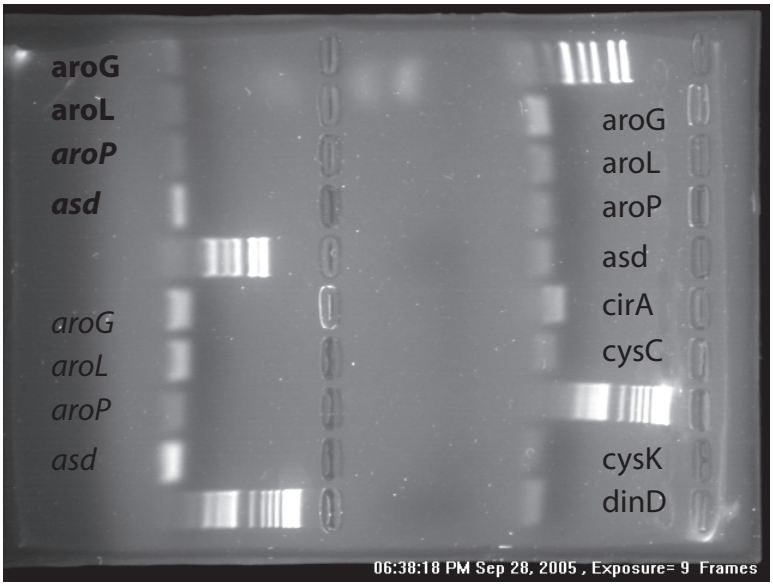


Figure 1.2: 80 ml, 1% agarose gel with 2.5 ul of 1% ethidium bromide run for 40 min at 100 volts. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used. Bold text indicates reaction was 100ul, italics represents a 3-step PCR. 6ul of each PCR was used. Product sizes for the 8 genes should be: aroG=173bp, aroL=161bp, aroP=122bp, asd=152bp, cirA=240bp, cysC=174bp, cysK=117, dinD=162bp. WARNING asd is probably designed for wrong region, need to check.

Brief Conclusions: It appears that all primer sets tested showed amplification of a single band of the correct length. However, the gel itself is very difficult to read. It wasn't run long enough and the short bp fragments are very fuzzy. Also the far left of the figure shows that migration of etBr towards the opposite pole of the gel is a problem when running short fragments.

1.1.2 Improving gel of target genes

Thu Sep 29 14:30:28 EDT 2005

In Figure 1.2, it appears that all primers worked fine, but the image/gel is bad. Using the same PCR products, we'll try a more appropriate 2% gel run longer with a post-stain to prevent the effects of having etBr migrate the opposite direction. Also going to try a larger comb and different concentrations of PCR product. 1x TAE was made fresh from the premade 50x Fisher stock in the Collins lab for both the gel and the running buffer.

For large (6-well) comb: trying aroG 6ul, 4ul, 3ul, cirA 6ul, 3ul
For medium (10-well) comb: trying aroG 6ul, 4ul, 3ul, cirA 6ul, 4ul, 3ul

Brief Conclusions: The gel in Figure 1.3 is much better than the previous one (Figure 1.2). The amplifications are definitely the correct size and all the bands of the ladder are clear. Adding 5 minutes to the post-stain and the post-water wash might help even more.

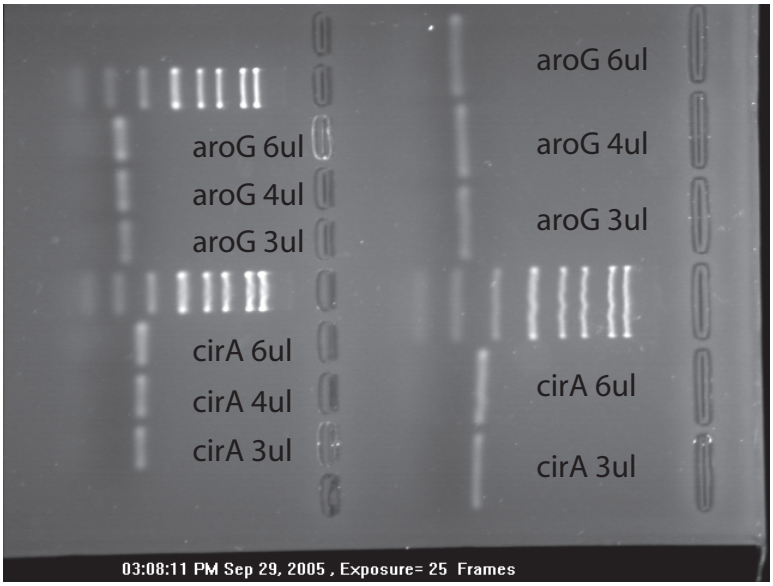


Figure 1.3: 70 ml, 2% agarose gel run for 40 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 40 min followed by 15 min in water alone (both on orbital shaker at 50rpm) to reduce background. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used. Product sizes for the 2 genes should be: aroG=173bp, cirA=240bp.

1.1.3 Higher-throughput primer organization

Fri Sept 30, 2005

Previously primers were in 1.5ml eppy tubes premixed with a forward and reverse primer for an intergenic sequence of a particular gene. To speed up primer testing and more importantly improve accuracy when testing genes on a 384-well qPCR plate the primers have been moved to a 96-well Costar plate. The concentration of each is 4μM allowing use of a 800pM final concentration in a 10 or 15μl reaction using 2 or 3μl respectively (the bottom precision range of our multichannel pipettor).

The primer organization is shown in Table 1.1.3. Each transcription factor (TF) is tested by one or more rows (indicated in the far left column; the first three columns are negative controls (genes thought not to be regulated by the transcription factor); the next two columns are positive controls (known targets of the TF). For example the first row tests lexA with negative controls: serA, entC, and fliF; positive controls: recA and lexA; and potential new targets: dinD, dinG, dinI, dinP, ruvA, yceP, yebG.

1.1.4 Higher-throughput primer check

Mon Oct 3 14:01:38 EDT 2005

Primers from the plate in Table 1.1.3 are to be tested as described in Section 1.1. All of the genes on the plate will be tested using the following reaction:

96-well primer plate for ChIP-PCR												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A: lexA	serA	entC	fliF	recA	lexA	dinD	dinG	dinI	dinP	ruvA	yceP	yebG
B: fliA	serA	entC	recA	fliF	fliL	nikA	flgK	flgM	fliC	-	-	-
C: fecI	serA	recA	fliF	fecA	fecI	cirA	entC	exbB	fepA	fhuA	nrdH	tonB
D: lrp	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB
E: lrp	ilvC	lysC	metA	metE	metF	dapB	dapD	-	-	-	-	-
F	-	-	-	-	-	-	-	-	-	-	-	-
G	-	-	-	-	-	-	-	-	-	-	-	-
H	-	-	-	-	-	-	-	-	-	-	-	-

Table 1.1: Primer organization. TFs tested by each row are indicated in the leftmost column. Hyphens indicate wells with H_2O only.

PCR Reaction composition	
H_2O	15 μl
Qiagen Master Mix	25 μl
Forward and reverse primer	10 μl
final primer concentration	900 pM

Thermal cycler conditions		
Initial denaturation	5 min	95°C
2-Step cycling		
Denature:	30 sec	95°C
Anneal/Extend:	45 sec	60°C
Number of Cycles:	25	
Final Extention:	7 min	72°C

Reactions are run in 4, 8-tube PCR strips (A5-A12, B2-B9, D5-D12, E1-E7) and one 12-tube strip (C1-C12) (actually one 8-tube and one 8-tube ripped in half).

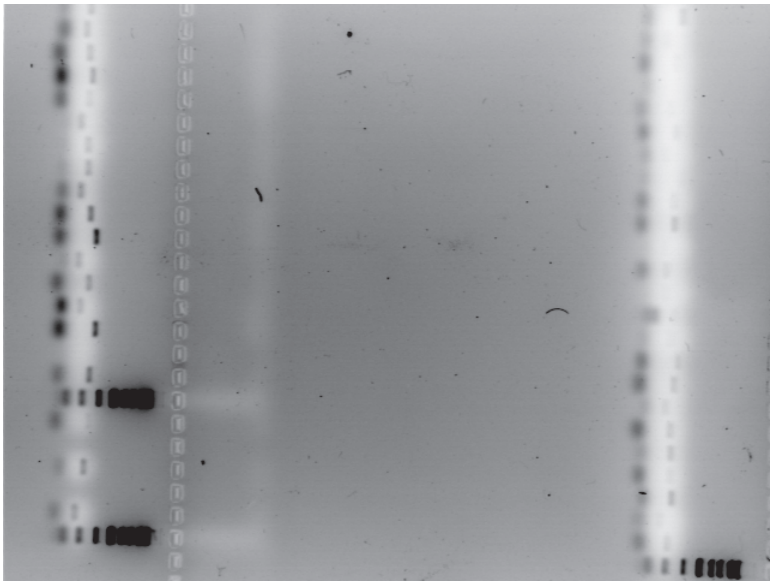


Figure 1.4: 200 ml, 2% agarose gel with 4 ul of 1% ethidium bromide run for 45 min at 120 volts. Image is the inverse as the large amount of dye on the gel made the true image very difficult to see. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used.

Brief Conclusions: A 250ml or 300ml gel would definitely be easier! The 200ml gel was very thin. Also, as can be seen, in 1.4 there is way too much loading dye and it migrates to the same

region as most of the PCR fragments. Last the imaging system used is the one from the CAB. The picture is of so low resolution that it isn't possible to make out any details.

1.1.5 Higher-throughput primer check: trying for better gels

Tue Oct 4 12:57:55 EDT 2005

Going to try running the lower-throughput 10-well gels with the post stain as was optimized in section 1.1.2. Samples will be prepared as before using multichannel pipettors, but will be loaded with a single-channel pipettor, as the gel-wells don't line up.

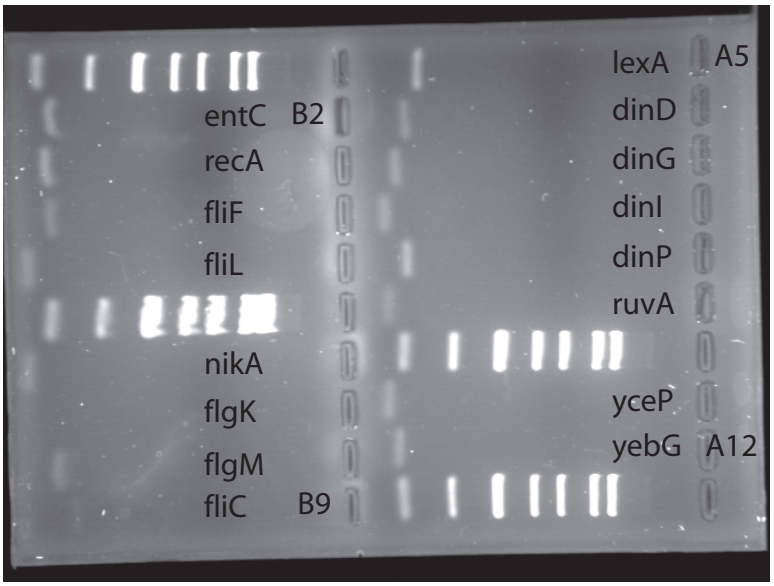


Figure 1.5: 70 ml, 2% agarose gel run for 45 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 45 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background Product sizes for the genes should be: lexA: 197, dinD: 162, dinG: 128, dinI: 101, dinP: 201, ruvA: 106, yceP: 111, yebG: 141, entC, recA: 153, fliF: 156, fliL: 151, nikA: 101, flgK: 150, flgM: 165, fliC: 216. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used.

Brief Conclusions: All in all the results aren't too bad. Annoyingly, there is a small amount of signal around the 30-50bp region that looks like primers are amplifying nonspecific regions as well as the correct one. This could be the result of many things including:

1. it's an artifact the 2-step PCR
2. Tm needs to be raised because primer3 incorrectly estimates melting temperature
3. bad primers
4. amplification from whole cells is messy

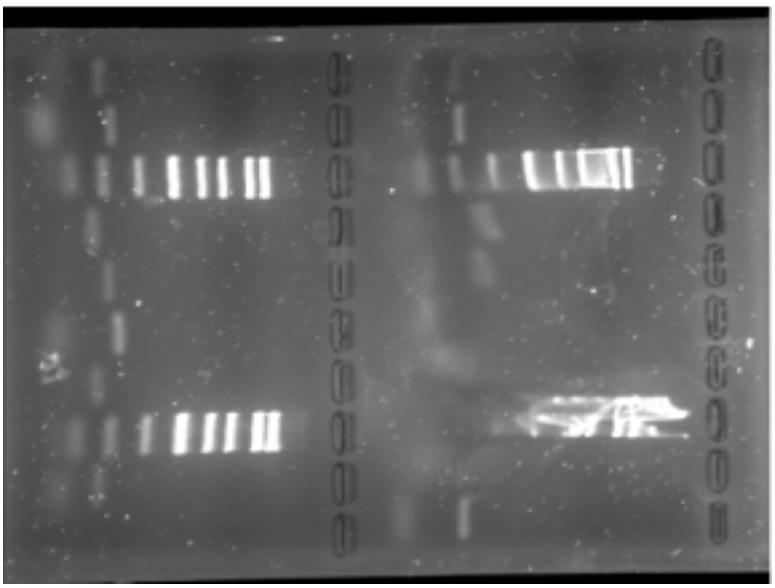


Figure 1.6: 70 ml, 2% agarose gel run for 38 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 45 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. This gel sucks, wasn't poured smoothly or melted properly. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used.

Hopefully, the last is correct. For sure dumping a little overnite culture into your PCR master mix isn't as good as using pure DNA to amplify from. We should try a subset of the genes again on the cleaned up, sheared whole-cell extract before immunoprecipation to see if they are likely to influence our qPCR results (which uses a 2-step process). fhuA, flgK, fliC, and cysC should be repeated to see if their faint bands are the result of pipetting error or maybe those primers are no good. The aroP fragment looks too short and should be run again as well.

1.2 Testing the shearing range

It is desireable to have the range of the chromatin-fixed, sheared DNA to be between 200 and 1000bp. It is also important to know the approximate amount of pre-precipitation starting DNA. To test this we need to perform the initial steps ChIP reaction:

1. from a 1:50 dilution grow cells in a flask to OD 0.6.
2. fix cells with 1% formaldehyde solution
3. wash cells 2x with PBS (to remove formaldehyde)
4. lyse cells (with lysozyme or readylyse)
5. shear DNA by sonication

Typically the above precedure is followed by immunoprecipitation. But we can take the lysate, reverse the crosslinks, and do a phenol:chloroform purification and ethanol precipitation to quantify

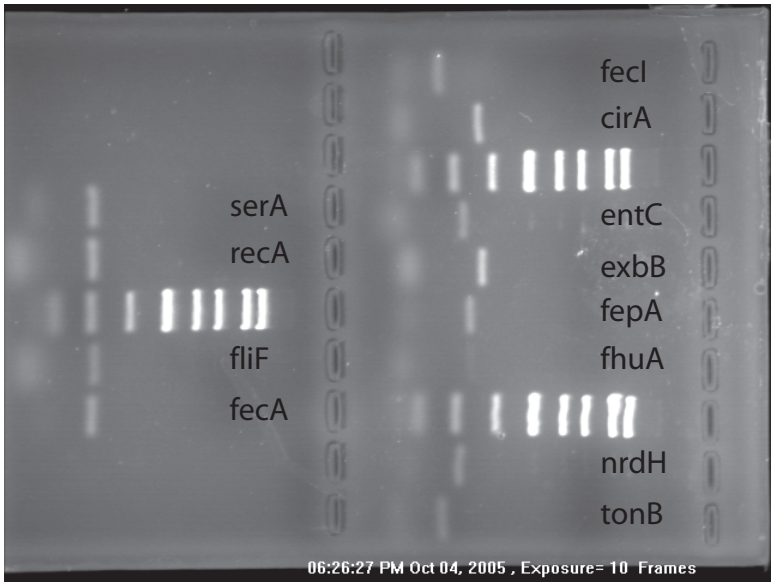


Figure 1.7: 70 ml, 2% agarose gel run for 37 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 45 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. Product sizes for the genes should be: fecI: 111, cirA: 240, entC: 175, exbB: 241, fepA: 195, fhuA: 199, nrdH: 162, tonB: 123, serA: 159, recA: 153, fliF: 156, fecA: 153. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used.

the amount of starting DNA and the size range. It is probably also a good idea to preceed the phenol:choloroform reaction with an RNAase digestion or else the gel will contain RNA too making it difficult to determine the size range.

1.2.1 First shear test

Tue Oct 11, 2005

1. Growing two strains (lexA:A, and lrp-myc:B) in a 50ml of LB in a 250ml culture flask and a 250ml baffled flask (4 samples total). Cultures started with a 1:100 dilution from an overnite culture. All cultures contain 50 μl Ampicillin (100mg/ml stock solution).

2. After 1hr, 25 μl IPTG (1M stock solution) was added to each culture to induce production of the transcription factor

3. After an additional 2hr 15min, two 10ml samples were taken from each flask and placed into 15ml Falcon centrifuge tubes for crosslinking (8 samples total named lexN1, lexN2, etc using the basename below). Also, 300 μl was taken from each culture to take an OD reading:

Strain	**Flask**	**OD 600**	**basename**
lexA:A	culture	0.26	lexN
lexA:A	baffled	0.27	lexB
lrp-myc:B	culture	0.22	lrpN
lrp-myc:B	baffled	0.18	lrpB

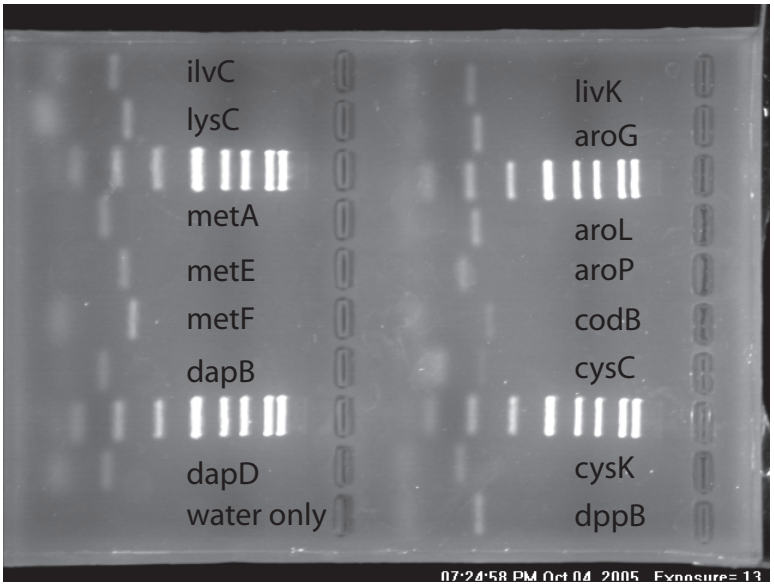


Figure 1.8: 70 ml, 2% agarose gel run for 34 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 45 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. Product sizes for the genes should be: livK: 152, aroG: 173, aroL: 161, aroP: 122, codB: 206, cysC: 174, cysK: 117, dppB: 170, ilvC: 132, lysC: 180, metA: 104, metE: 158, metF: 187, dapB: 101, dapD: 117. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used.

4. 280 μl of 37% formaldehyde was added to each 10ml sample and mixed by inversion. Sample was incubated at RT for 10min
5. cells were pelleted at 2900g for 8 min and washed 2x with PBS
6. washed pellets were lysed for 30min at 37C in 500 μl of Pallson lysis buffer (no-shaking)
7. 500 μl of 2x Pallson IP buffer with 1 μl of RNAaseA (1 μg/μl stock [Ambion])was added and the tubes were placed in a shaking incubator (300rpm) for 10min yielding a very clear lysate
8. the 1ml lysate was transferred to a 1.5ml eppy tube
9. lysed samples were sonicated with a Branson Sonicator at 3 x 20 sec with power 1.5; samples lexB2 and lrpB2 (the second samples from the baffled flask) were sonicated a 4th time
10. sonicated lysates were spun at top speed (13,200 rpm) for 5 min at 4C to pellet cellular debris (there was *very* little)
11. supernants were transferred to new eppy tubes
12. 2.5 μl of Proteinase K (20 μg/μl stock [Ambion]) was added to each tube and they were put in a 65C heat-block overnite to remove cross-links

ERROR: Palsson 2x IP buffer was made with 8% Triton X-100 not the correct 4%

More precise times:

Tue Oct 11 11:57:33 EDT 2005 in incubator

Tue Oct 11 12:50 EDT 2005 added 25mM IPTG

Tue Oct 11 2:15 EDT 2005 out incubator

Tue Oct 11 19:11:20 EDT 2005 put sheared chromatin-DNA in 65C heat-block to remove cross-links.

Wed Oct 12 10:42:47 EDT 2005 removed from heat-block

Wed Oct 12 14:28:55 EDT 2005 removed from -85C freezer to centrifuge

Wed Oct 12 18:43:51 EDT 2005 running gel with sheared DNA to check size range

DNA was phenol:chloroform extracted and ethanol precipitated using the method of Barker. The Ethanol/DNA 2:1 mix was placed in the -85C for 60min followed by centrifugation for 20 min at 4C.

DNA quantification

Thu Oct 12, 2005

Each of the eight samples was quantified using the NanoDrop. Different amounts of lysate (200μl or 100μl of the total 1ml) were used to see if things scaled linearly. The sample nomenclature is as follows [samplename : amount of lysate used for extraction].

Sample	**DNA (ng/μl)**	**260/280**	**260/230**	**total DNA (ug)** [1]
lexN1:200	364.0	2.06	2.28	182.0
lexN2:200	327.5	2.05	2.31	163.8
lexN2:100	134.9	2.06	2.27	134.9
lexB1:200	283.1	2.08	2.30	141.6
lexB2:200	285.1	2.05	2.31	142.6
lrpN1:200	203.4	2.06	2.31	101.7
lrpN2:200	205.3	2.08	2.32	102.7
lrpB1:200	135.6	2.03	2.23	67.8
lrpB2:200	156.9	2.05	2.21	78.5
lrpB2:100	74.5	2.05	2.19	74.5

Brief Conclusions: It looks like the sheared DNA yield is a function of the cell density of the culture it was taken from (see Figure 1.9) [not surprising] and better yields are obtained when extracting larger fractions of culture (probably because it's easier to get more of the aqueous from the phenol:aqueous interface).

DNA shearing range

1μl and 2μl of each samples lexN2:200, lexN2:100, lrpB2:200, and lrpB2:100 were run on an agarose gel.

Brief Conclusions: It looks like the shearing range is too high. This result could be due to one of two problems: 1 we didn't shear long enough or on high enough power, 2 there is a lot of RNA masking the DNA signal. We can do an RNAse digest on the DNA to see if that cleans it up to test hypothesis 2.

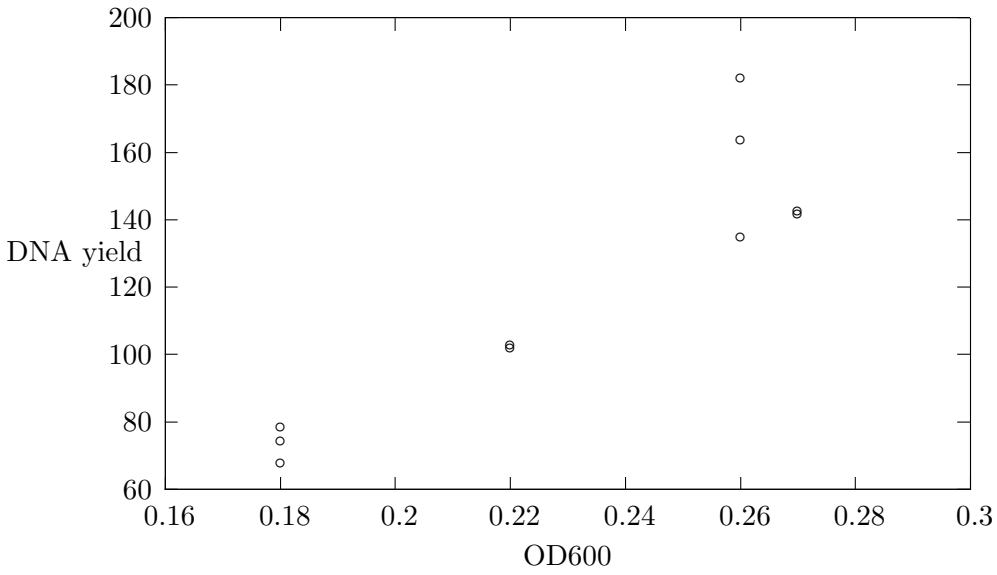


Figure 1.9: amount of sheared DNA obtained from lysate as a function of the OD of the culture it was extracted from.

RNAse digesting sheared DNA

Thu Oct 13 18:37:19 EDT 2005 lexAN2:200 and lrpB2:200 will be RNAse digested to remove any contaminating RNA.

The RNAseA / DNA mixture was purified using a Qiagen PCR cleanup kit. DNA was resuspended in 50μl , roughly half the original volume. Cleaned up yields were:

Sample	**DNA (ng/μl)**	**260/280**	**260/230**
lexAN2:100	55.8	1.90	2.16
lrpB2:200	49.3	1.86	2.10

Brief Conclusions: The Nanodrop spec shows a cleaner DNA as the 260/280 is closer to the more ideal 1.8, but the yield is about half its pre-RNAse digested value (so either there was a LOT of RNA or the Qiagen columns are very inefficient). Also, the gel from the RNAse digested samples looks cleaner, but the sheared range is still a little on the high side. Perhaps the lrpB2, which was sheared 20 extra seconds is a little smaller average size but not that much. Next round we should do at least that much shearing. Also, the RNAse digest was done with an Ambion RNAse A enzyme. I have purchase an Ambion RNAse cocktail that should chop up the RNA to a smaller size range next time in case there was a little left. Not sure why the ladder looks bad. I changed out the buffer for the next gel in case it was that.

Primer testing on sheared DNA

PCR: Thur Oct 13, 2005

Doing a quick positive control / sanity check to make sure the primers tested earlier can amplify the sheared DNA. Sheared DNA was from RNAse digested lrpB2:200. 1.3μg was used to amplify

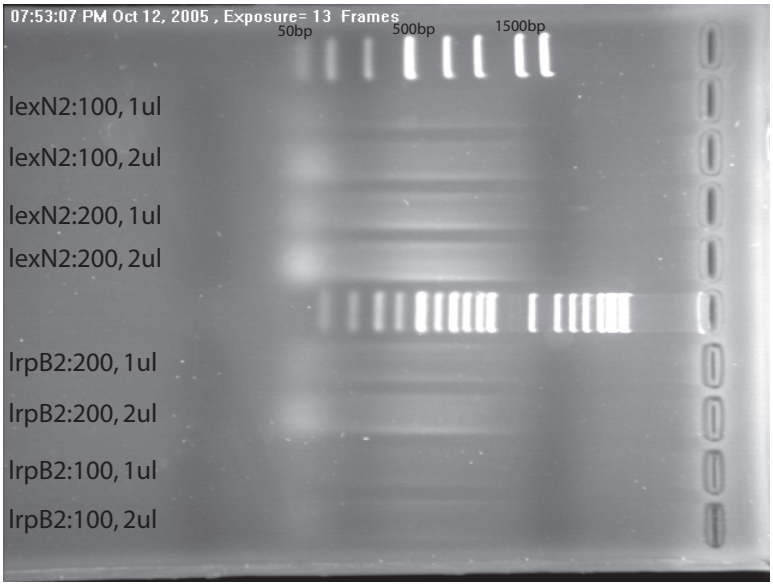


Figure 1.10: 70 ml, 1.5% agarose gel run for 48 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 32 min followed by 10 min in water alone (both on orbital shaker at 50rpm) to reduce background. 10 μl of Fisher BioReagents exACTGene 1KB Plus DNA Ladder, with bands:weights(ng) of 10000:43, 8000:31, 6000:28, 5000:41, 4000:36, 3000:31, 2500:26, 2000:23, 1500:22, 1000:35, 900:26, 800:52, 700:31, 600:18, 500:35, 400:12, 300:17, 200:12, 100:7 was used. Desired shearing range is 200-300bp.

all 44 tested targets (I just pipetted it into the multichannel reservior).

Gel: Fri Oct 14 19:05:27 EDT 2005

Brief Conclusions: It is clear when comparing Figure 1.12 to those from section 1.1.3 that the PCR from cleaned up sheared DNA is much more effecient and cleaner than when I dumped in 10μl of overnite to amplify from. The only question this figure raises is: What happened with codB? Earlier cysC was tentative, but here this gene amplifies cleanly. Now codB doesn't show up at all. The best part of this figure is there are no fuzzy bands around the 50bp point as was consistently found when amplifying from whole cells. Perhaps sheared DNA should be used when cloning in the any future TOPO constructs? Also, it should be noted that this DNA was cleaned 2x, once with phenol:chloroform extraction and once by a Qiagen PCR purification kit.

1.3 Checking the cloned TOPO constructs

Two Invitrogen TOPO vectors were used for cloning transcription factors to add tags to the C terminal or N terminal. The pTrcHis vector adds N-terminal 6xHis and Xpress epitopes. While the pTrcHis2 add C-terminal c-myc and 6xHis epitopes. We must sequence the vectors to confirm there are no PCR errors and that the gene is inserted in frame. All of the vectors were previously verified by restriction digest, but the digest will be repeated prior to sequencing to make certain all freezer stocks were labeled properly. Sequences will be send with primer and template mixed.

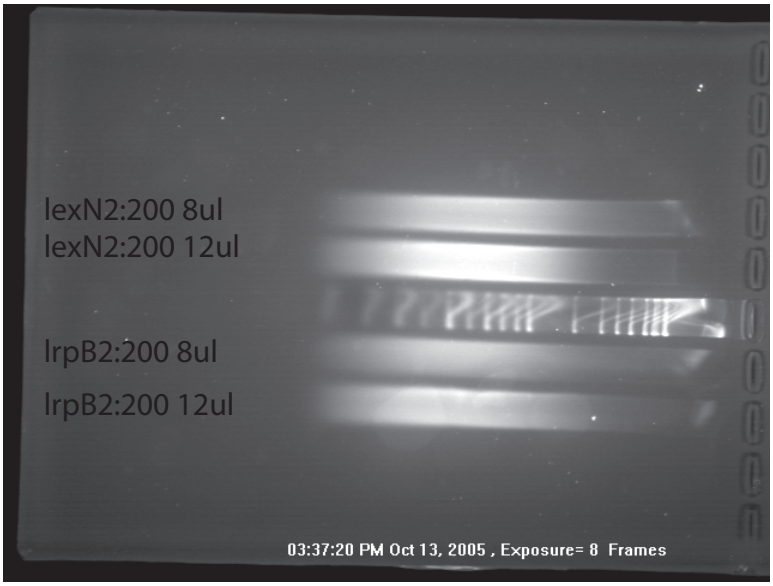


Figure 1.11: 70 ml, 1.5% agarose gel run for 48 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 32 min followed by 10 min in water alone (both on orbital shaker at 50rpm) to reduce background. 10 μl of Fisher BioReagents exACTGene 1KB Plus DNA Ladder, with bands:weights(ng) of 10000:43, 8000:31, 6000:28, 5000:41, 4000:36, 3000:31, 2500:26, 2000:23, 1500:22, 1000:35, 900:26, 800:52, 700:31, 600:18, 500:35, 400:12, 300:17, 200:12, 100:7 was used. Desired shearing range is 200-300bp.

1.3.1 DNA Sequencing

Sequencing is to be done using the Agencourt QuickLane Sequencing service. Required concentrations are:

Amount of Template	**Amount of Primer**	**Total Volume**
0.6-1.0ug	20pmol	40μl

The concentration of primer roughly corresponds to 1μl of the primers supplied with the Invitrogen kit. For the pTrcHis vector we are using the Xpress forward primer. For the pTrcHis2 sequences we are using the pTrcHis Reverse priming site.

Minipreps were performed using an eppendorf kit on overnite cultures from the following plasmids: fec-myc, lrp-myc, lrp, fliA, lexA. Yields were determined with the Nanodrop and are as follows (additionally the amount of plasmid needed for a 600ng DNA sequencing reaction are included):

Gene	**Amount of Vector (ng/μl)**	**260/280**	**260/230**	**μl for 600 ng**
fecI-myc	32.3	2.12	0.30	18.57
lrp-myc	31.3	2.03	0.35	19.17
lrp	36.7	2.02	0.29	16.35
fliA	26.3	2.03	0.38	22.81
lexA	39.9	2.22	0.19	15.04

Restriction digests to reconfirm insert size are being run before sending the constructs for sequencing. The following reactions are being performed:

lrp, fliA, lexA

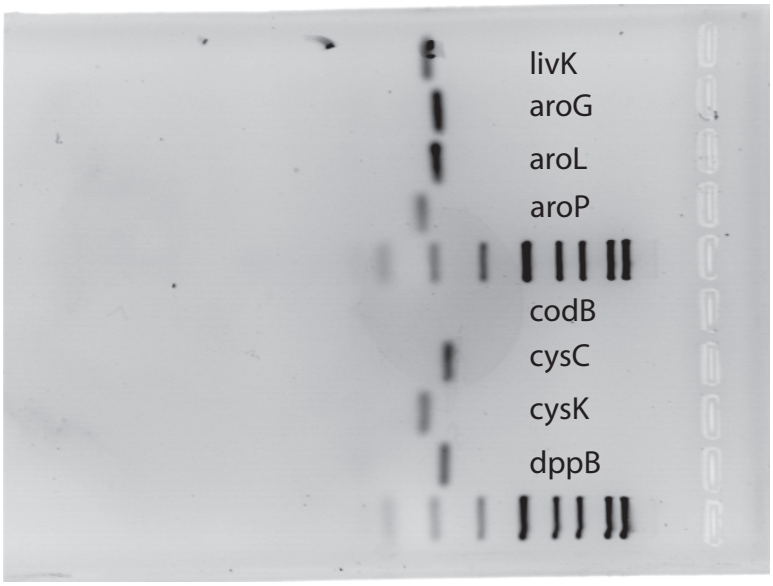


Figure 1.12: 70 ml, 2% agarose gel run for 38 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 40 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used. Product sizes for the genes should be: livK: 152, aroG: 173, aroL: 161, aroP: 122, codB: 206, cysC: 174, cysK: 117, dppB: 170.

Restriction Digest	
vector/DNA	9 μl
EcoRI buffer	2 μl
BSA	2 μl
EcoRI enzyme	0.5 μl
BamHI enzyme	0.5 μl
H_2O	6 μl

lrp-myc, fecI-myc

Restriction Digest	
vector/DNA	9 μl
EcoRI buffer	2 μl
BSA	0 μl
EcoRI enzyme	0.5 μl
NcoI enzyme	0.5 μl
H_2O	8 μl

Digests placed in 37C incubator at: Mon Oct 10 15:07:14 EDT 2005

out at: Mon Oct 10 16:07 EDT 2005

As seen in Figure 1.13, the inserts are all the appropriate sizes and will be sent for sequencing.

Vectors were picked up by Agencourt courier for sequencing on: Tue Oct 11 11:27:49 EDT 2005 with project name: *ChIP TF clones*

4 of 5 sequences passed the Agencourt QC (800bp Phred 20), fecI didn't and is being resequenced

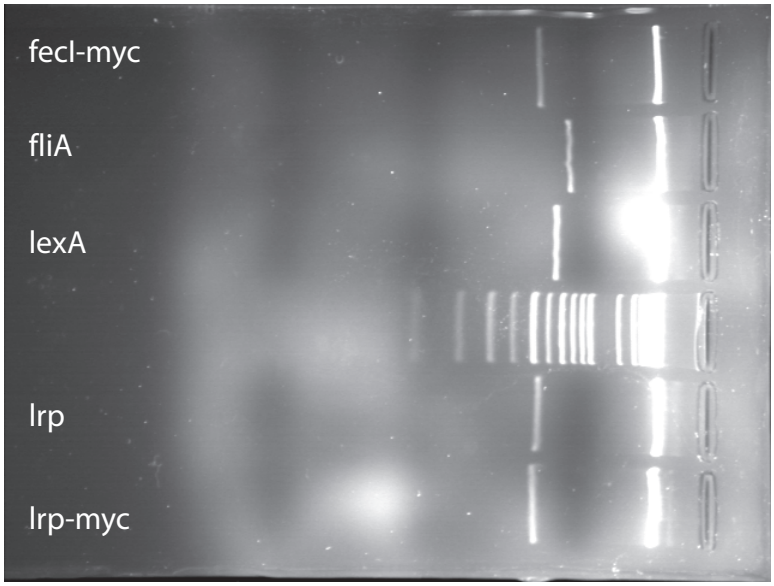


Figure 1.13: 70 ml, 2% agarose gel with 0.75 ul of 1% ethidium bromide run for 40 min at 120 volts. 10 μl of Fisher BioReagents exACTGene 1KB Plus DNA Ladder, with bands:weights(ng) of 10000:43, 8000:31, 6000:28, 5000:41, 4000:36, 3000:31, 2500:26, 2000:23, 1500:22, 1000:35, 900:26, 800:52, 700:31, 600:18, 500:35, 400:12, 300:17, 200:12, 100:7 was used. 20ul (all) of each digest was used. Product sizes for the 8 genes should be: fecI-myc=521bp, fliA=719bp, lexA=608bp, lrp=494bp, lrp-myc=494bp.

```
LUSTAL W (1.82) multiple sequence alignment


lrp                             --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq    --------------------------------------------------
lrp_TOPO_pTrcHis2.trimmed.seq   CAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGGTCGTAAATCACTG 50


lrp                             --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq    --------------------------------------------------
lrp_TOPO_pTrcHis2.trimmed.seq   CATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTGC 100


lrp                             --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq    --------------------------------------------------
lrp_TOPO_pTrcHis2.trimmed.seq   GCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAA 150


lrp                             --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq    --------------------------------------------------
lrp_TOPO_pTrcHis2.trimmed.seq   TTAATCATCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTT 200


lrp                             --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq    --------------------------------------------------
lrp_TOPO_pTrcHis2.trimmed.seq   CACACAGGAAACAGCGCCGCTGAGAAAAAGCGAAGCGGCACTGCTCTTTA 250


lrp                             --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq    --------------------------------------------------
lrp_TOPO_pTrcHis2.trimmed.seq   ACAATTTATCAGACAATCTGTGTGGGCACTCGACCGGAATTATCGATTAA 300


lrp                             --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq    --------------------------------------------------
lrp_TOPO_pTrcHis2.trimmed.seq   CTTTATTATTAAAAATTAAAGAGGTATATATTAATGTATCGATTAAATAA 350


lrp                             ----------------------ATGGTAGATAGCAAGAAGCGCCCTGGCA 28
lrp_TOPO_pTrcHis.trimmed.seq    -----------------CCCTTATGGTAGATAGCAAGAAGCGCCCTGGCA 33
lrp_TOPO_pTrcHis2.trimmed.seq   GGAGGAATAAACCATGGCCCTTATGGTAGATAGCAAGAAGCGCCCTGGCA 400
                                           ****************************

lrp                             AAGATCTCGACCGTATCGATCGTAACATTCTTAATGAGTTGCAAAAGGAT 78
lrp_TOPO_pTrcHis.trimmed.seq    AAGATCTCGACCGTATCGATCGTAACATTCTTAATGAGTTGCAAAAGGAT 83
```

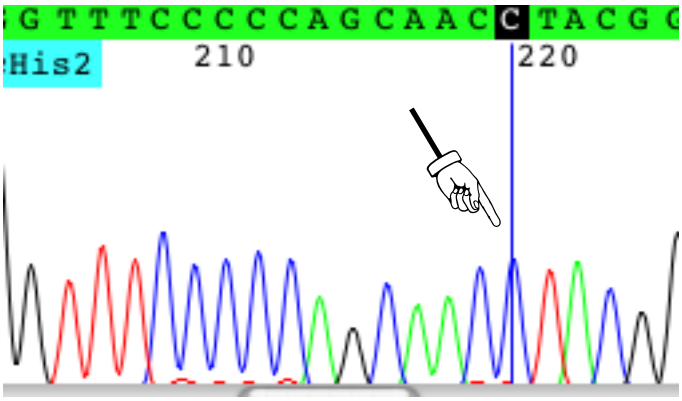


Figure 1.14: Where the finger points should be a **T**.

```
lrp_TOPO_pTrcHis2.trimmed.seq      AAGATCTCGACCGTATCGATCGTAACATTCTTAATGAGTTGCAAAAGGAT 450
                                   **************************************************

lrp                                GGGCGTATTTCTAACGTCGAGCTTTCTAAACGTGTGGGACTTTCCCCAAC 128
lrp_TOPO_pTrcHis.trimmed.seq       GGGCGTATTTCTAACGTCGAGCTTTCTAAACGTGTGGGACTTTCCCCAAC 133
lrp_TOPO_pTrcHis2.trimmed.seq      GGGCGTATTTCTAACGTCGAGCTTTCTAAACGTGTGGGACTTTCCCCAAC 500
                                   **************************************************

lrp                                GCCGTGCCTTGAGCGTGTGCGTCGGCTGGAAAGACAAGGGTTTATTCAGG 178
lrp_TOPO_pTrcHis.trimmed.seq       GCCGTGCCTTGAGCGTGTGCGTCGGCTGGAAAGACAAGGGTTTATTCAGG 183
lrp_TOPO_pTrcHis2.trimmed.seq      GCCGTGCCTTGAGCGTGTGCGTCGGCTGGAAAGACAAGGGTTTATTCAGG 550
                                   **************************************************

lrp                                GCTATACGGCGCTGCTTAACCCCCATTATCTGGATGCATCACTTCTGGTA 228
lrp_TOPO_pTrcHis.trimmed.seq       GCTATACGGCGCTGCTTAACCCCCATTATCTGGATGCATCACTTCTGGTA 233
lrp_TOPO_pTrcHis2.trimmed.seq      GCTATACGGCGCTGCTTAACCCCCATTATCTGGATGCATCACTTCTGGTA 600
                                   **************************************************

lrp                                TTCGTTGAGATTACTCTGAATCGTGGCGCACCGGATGTGTTTGAACAATT 278
lrp_TOPO_pTrcHis.trimmed.seq       TTCGTTGAGATTACTCTGAATCGTGGCGCACCGGATGTGTTTGAACAATT 283
lrp_TOPO_pTrcHis2.trimmed.seq      TTCGTTGAGATTACTCTGAATCGTGGCGCACCGGATGTGTTTGAACAATT 650
                                   **************************************************

lrp                                CAATACCGCTGTACAAAAACTTGAAGAAATTCAGGAGTGTCATTTAGTAT 328
lrp_TOPO_pTrcHis.trimmed.seq       CAATACCGCTGTACAAAAACTTGAAGAAATTCAGGAGTGTCATTTAGTAT 333
lrp_TOPO_pTrcHis2.trimmed.seq      CAATACCGCTGTACAAAAACTTGAAGAAATTCAGGAGTGTCATTTAGTAT 700
                                   **************************************************

lrp                                CCGGTGATTTCGACTACCTGTTGAAAACACGCGTGCCGGATATGTCAGCC 378
lrp_TOPO_pTrcHis.trimmed.seq       CCGGTGATTTCGACTACCTGTTGAAAACACGCGTGCCGGATATGTCAGCC 383
lrp_TOPO_pTrcHis2.trimmed.seq      CCGGTGATTTCGACTACCTGTTGAAAACACGCGTGCCGGATATGTCAGCC 750
                                   **************************************************

lrp                                TACCGTAAGTTGCTGGGGGAAACCCTGCTGCGTCTGCCTGGCGTCAATGA 428
lrp_TOPO_pTrcHis.trimmed.seq       TACCGTAAGTTGCTGGGGGAAACCCTGCTGCGTCTGCCTGGCGTCAATGA 433
lrp_TOPO_pTrcHis2.trimmed.seq      TACCGTAGGTTGCTGGGGGAAACCCTGCTGCGTCTGCCTGGCGTCAATGA 800
                                   ******* ****************************************

lrp                                CACACGGACATACGTTGTTATGGAAGAAGTCAAGCAGAGTAATCGTCTGG 478
lrp_TOPO_pTrcHis.trimmed.seq       CACACGGACATACGTTGTTATGGAAGAAGTCAAGCAGAGTAATCGTCTGG 483
lrp_TOPO_pTrcHis2.trimmed.seq      CACACGGACATACGTTGTTATGGAAGAAGTCAAGCAGAGTAATCGTCTGG 850
                                   **************************************************

lrp                                TTATTAAGACGCGCTAA---------------------------------- 495
lrp_TOPO_pTrcHis.trimmed.seq       TTATTAAGACGCGCTAAAAGGGCGAATTCGAAGCTTGGCTGTTTTGGCGG 533
lrp_TOPO_pTrcHis2.trimmed.seq      TTATTAAGACGCGC---AAGGGCGAATTCGAAGCTTACGTAGAACAAAAA 897
                                   **************

lrp                                --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq       ATGAG---AGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAA 580
lrp_TOPO_pTrcHis2.trimmed.seq      CTCATCTCAGAAGAGGATCTGAATAGCGCCG-TCGACCATCATCATCATC 946


lrp                                --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq       GCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCA 630
lrp_TOPO_pTrcHis2.trimmed.seq      ATCATTGAGTTAAACGGTCTCCAGCT------------------------ 972


lrp                                --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq       CCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAG 680
lrp_TOPO_pTrcHis2.trimmed.seq      --------------------------------------------------


lrp                                --------------------------------------------------
```

```
lrp_TOPO_pTrcHis.trimmed.seq       TGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAA 730
lrp_TOPO_pTrcHis2.trimmed.seq      --------------------------------------------------


lrp                                --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq       CGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTC 780
lrp_TOPO_pTrcHis2.trimmed.seq      --------------------------------------------------


lrp                                --------------------------------------------------
lrp_TOPO_pTrcHis.trimmed.seq       GGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACG 830
lrp_TOPO_pTrcHis2.trimmed.seq      --------------------------------------------------


lrp                                ------
lrp_TOPO_pTrcHis.trimmed.seq       TTGCGA 836
lrp_TOPO_pTrcHis2.trimmed.seq      ------


CLUSTAL W (1.82) multiple sequence alignment

fliA is reverse complemented

fliA_TOPO_pTrcHis.trimmed.seq      CGCATGGGGAGACCCCACACTACCATCGGCGCTACGGCGTTTCACTTCTG 50
EG11355                            --------------------------------------------------


fliA_TOPO_pTrcHis.trimmed.seq      AGTTCGGCATGGGGTCAGGTGGGACCACCGCGCTACTGCCGCCAGGCAAA 100
EG11355                            --------------------------------------------------


fliA_TOPO_pTrcHis.trimmed.seq      TTCTGTTTTATCAGACCGCTTCTGCGTTCTGATTTAATCTGTATCAGGCT 150
EG11355                            --------------------------------------------------


fliA_TOPO_pTrcHis.trimmed.seq      GAAAATCTTCTCTCATCCGCCAAAACAGCCAAGCTTCGAATTCGCCCTTG 200
EG11355                            -------------------------------------------------G 1
                                                                                     *

fliA_TOPO_pTrcHis.trimmed.seq      TGAATTCACTCTATACCGCTGAAGGTGTAATGGATAAACACTCGCTGTGG 250
EG11355                            TGAATTCACTCTATACCGCTGAAGGTGTAATGGATAAACACTCGCTGTGG 51
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      CAGCGTTATGTCCCGCTGGTGCGTCACGAAGCATTGCGCCTGCAGGTTCC 300
EG11355                            CAGCGTTATGTCCCGCTGGTGCGTCACGAAGCATTGCGCCTGCAGGTTCG 101
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      ACTGCCCGCGAGCGTGGAACTTGACGATCTGCTACAGGCGGGCGGCATTG 350
EG11355                            ACTGCCCGCGAGCGTGGAACTTGACGATCTGCTACAGGCGGGCGGCATTG 151
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      GGTTACTTAATGCCGTCGAACGCTATGACGCCCTACAAGGAACGGCATTT 400
EG11355                            GGTTACTTAATGCCGTCGAACGCTATGACGCCCTACAAGGAACGGCATTT 201
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      ACAACTTACGCAGTGCAGCGTATCCGTGGCGCTATGCTGGATGAACTTTG 450
EG11355                            ACAACTTACGCAGTGCAGCGTATCCGTGGCGCTATGCTGGATGAACTTCG 251
                                   ********************************************** *

fliA_TOPO_pTrcHis.trimmed.seq      CAGCCGTGACTGGGTGCCGCGCAGCGTGCGACGCAACGCGCGTGAAGTGG 500
EG11355                            CAGCCGTGACTGGGTGCCGCGCAGCGTGCGACGCAACGCGCGTGAAGTGG 301
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      CACAGGCAATAGGGCAACTGGAGCAGGAACTTGGCCGCAACGCCACGGAA 550
EG11355                            CACAGGCAATAGGGCAACTGGAGCAGGAACTTGGCCGCAACGCCACGGAA 351
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      ACTGAGGTAGCGGAACGTTTAGGGATCGATATTGCCGATTATCGCCAAAT 600
EG11355                            ACTGAGGTAGCGGAACGTTTAGGGATCGATATTGCCGATTATCGCCAAAT 401
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      GTTGCTCGACACCAATAACAGCCAGCTCTTCTCCTACGATGAGTGGCGCG 650
EG11355                            GTTGCTCGACACCAATAACAGCCAGCTCTTCTCCTACGATGAGTGGCGCG 451
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      AAGAGCACGGCGATAGCATCGAACTGGTTACTGATGATCATCAGCGAGAA 700
EG11355                            AAGAGCACGGCGATAGCATCGAACTGGTTACTGATGATCATCAGCGAGAA 501
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      AACCCGCTACAACAACTACTGGACAGTAATCTGCGCCAGCGGGTGATGGA 750
EG11355                            AACCCGCTACAACAACTACTGGACAGTAATCTGCGCCAGCGGGTGATGGA 551
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      AGCCATCGAAACGTTGCCGGAGCGCGAAAAACTGGTATTAACCCTCTATT 800
EG11355                            AGCCATCGAAACGTTGCCGGAGCGCGAAAAACTGGTATTAACCCTCTATT 601
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      ACCAGGAAGAGCTGAATCTCAAAGAGATTGGCGCGGTGCTGGAGGTCGGG 850
EG11355                            ACCAGGAAGAGCTGAATCTCAAAGAGATTGGCGCGGTGCTGGAGGTCGGG 651
                                   **************************************************

fliA_TOPO_pTrcHis.trimmed.seq      GAATCGCGGGTCAGTCAGTTACACAGCCAGGCTATTAAACAGTTGCGCAC 900
```

```
EG11355                         GAATCGCGGGTCAGTCAGTTACACAGCCAGGCTATTAAACGGTTACGCAC 701
                                **************************************** *** *****

fliA_TOPO_pTrcHis.trimmed.seq   TAAACTGGGTAAGTTATAAAAGGGTTGGATCCTAT 935
EG11355                         TAAACTGGGTAAGTTATAA---------------- 720


CLUSTAL W (1.82) multiple sequence alignment


lexA_TOPO_pTrcHis.trimmed.seq   GGATCCACCCTTATGAAAGCGTTAACGGCCAGGCAACAAGAGGTGTTTGA 50
EG10533                         ------------ATGAAAGCGTTAACGGCCAGGCAACAAGAGGTGTTTGA 38
                                            ************************************

lexA_TOPO_pTrcHis.trimmed.seq   TCTCATCCGTGATCACATCAGCCAGACAGGTATGCCGCCGACGCGTGCGG 100
EG10533                         TCTCATCCGTGATCACATCAGCCAGACAGGTATGCCGCCGACGCGTGCGG 88
                                **************************************************

lexA_TOPO_pTrcHis.trimmed.seq   AAATCGCGCAGCGTTTGGGGTTCCGTTCCCCAAACGCGGCTGAAGAACAT 150
EG10533                         AAATCGCGCAGCGTTTGGGGTTCCGTTCCCCAAACGCGGCTGAAGAACAT 138
                                **************************************************

lexA_TOPO_pTrcHis.trimmed.seq   CTGAAGGCGCTGGCACGCAAAGGCGTTATTGAAATTGTTTCCGGCGCATC 200
EG10533                         CTGAAGGCGCTGGCACGCAAAGGCGTTATTGAAATTGTTTCCGGCGCATC 188
                                **************************************************

lexA_TOPO_pTrcHis.trimmed.seq   ACGCGGGATTCGCCTGTTGCAGGAAGAGGAAGAAGGGTTGCCGCTGGTAG 250
EG10533                         ACGCGGGATTCGTCTGTTGCAGGAAGAGGAAGAAGGGTTGCCGCTGGTAG 238
                                ************ ***************************************

lexA_TOPO_pTrcHis.trimmed.seq   GTCGTGTGGCTGCCGGTGAACCACTTCTGGCGCAACAGCATATTGAAGGT 300
EG10533                         GTCGTGTGGCTGCCGGTGAACCACTTCTGGCGCAACAGCATATTGAAGGT 288
                                **************************************************

lexA_TOPO_pTrcHis.trimmed.seq   CATTATCAGGTCGATCCTTCCTTATTCAAGCCGAATGCTGATTTCCTGCT 350
EG10533                         CATTATCAGGTCGATCCTTCCTTATTCAAGCCGAATGCTGATTTCCTGCT 338
                                **************************************************

lexA_TOPO_pTrcHis.trimmed.seq   GCGCGTCAGCGGGATGTCGATGAAAGATATCGGCATTATGGATGGTGACT 400
EG10533                         GCGCGTCAGCGGGATGTCGATGAAAGATATCGGCATTATGGATGGTGACT 388
                                **************************************************

lexA_TOPO_pTrcHis.trimmed.seq   TGCTGGCAGTGCATAAAACTCAGGATGTACGTAACGGTCAGGTCGTTGTC 450
EG10533                         TGCTGGCAGTGCATAAAACTCAGGATGTACGTAACGGTCAGGTCGTTGTC 438
                                **************************************************

lexA_TOPO_pTrcHis.trimmed.seq   GCACGTATTGATGACGAAGTTACCGTTAAGCGCCTGAAAAAACAGGGCAA 500
EG10533                         GCACGTATTGATGACGAAGTTACCGTTAAGCGCCTGAAAAAACAGGGCAA 488
                                **************************************************

lexA_TOPO_pTrcHis.trimmed.seq   TAAAGTCGAACTGTTGCCAGAAAATAGCGAGTTTAAACCAATTGTCGTTG 550
EG10533                         TAAAGTCGAACTGTTGCCAGAAAATAGCGAGTTTAAACCAATTGTCGTTG 538
                                **************************************************

lexA_TOPO_pTrcHis.trimmed.seq   ACCTTCGTCAGCAGAGCTTCACCATTGAAGGGCTGGCGGTTGGGGTTATT 600
EG10533                         ACCTTCGTCAGCAGAGCTTCACCATTGAAGGGCTGGCGGTTGGGGTTATT 588
                                **************************************************

lexA_TOPO_pTrcHis.trimmed.seq   CGCAACGGCGACAAGGGCGAATTCGAAGCTTGGCTGTTTTGGCGGATGAG 650
EG10533                         CGCAACGGCGACTGGCTGTAA----------------------------- 609
                                ************  *    **

lexA_TOPO_pTrcHis.trimmed.seq   AGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTG 700
EG10533                         -------------------------------------------------

lexA_TOPO_pTrcHis.trimmed.seq   ATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCC 750
EG10533                         -------------------------------------------------

lexA_TOPO_pTrcHis.trimmed.seq   CATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGT 800
EG10533                         -------------------------------------------------

lexA_TOPO_pTrcHis.trimmed.seq   CTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGC 850
EG10533                         -------------------------------------------------

lexA_TOPO_pTrcHis.trimmed.seq   TCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACG 900
EG10533                         -------------------------------------------------

lexA_TOPO_pTrcHis.trimmed.seq   CTCTCCTG 908
EG10533                         -------- 
```

Notice the error at bp 758!

Brief Conclusions: As can be seen from the unfortunate sequences above, there are *MANY* errors. It looks like I need to use a proofreading Taq, I ordered *Easy-A High-Fidelity PCR Cloning Enzyme and Master Mix* designed specially for TOPO vectors because it adds the extra A on the end and uses a Pfu proofreading Taq. Supposedly has an error rate of less than 1 in 10 vectors. Hopefully, that will fix this problem.

One sequence *lrp* on the pTrcHis vector is ok. The fliA vector was inserted backwards. The other lrp sequence on the pTrcHis2 strain only has one mutation which converts a lysine to an arginine. These are similar amino acids both with positively charged R-groups so it might not matter, and it'd be interesting to see if it does.

The lexA vector is *almost* fine; there are no PCR errors, but unfortunately there is a human error. I removed the LAST THREE CODONS from the primer AHHHHHHHHHH!!!!! I hate human errors.

Last, something is screwy with the MiniPrep kit from eppendorf, I've been getting very low yields (30ng/μl). I will try the Qiagen kit, but last time I used this kit the results had a white precipitate that bothered me.

This weekend I need to fix the miniprep problem (trying both eppy and qiagen) so I can send all the vectors I've cloned out to see if there is anything useful in there before I start cloning.

1.3.2 Do any strains not have errors?

Sun Oct 16 14:51:03 EDT 2005

Only one in five clones from the previous round was error-free (though lexA was caused by a human error and one of the lrp's only have a single mutation. I have many previously cloned genes left, so I'm minipreping all I can find to try and see if any are usable and how much more cloning I need to perform.

gadX:A, gadX-myc:A, fecI:A, pdhR:A, pdhR-myc:A, flhC:A, yheO:A, sfsA:A, yhiF:A, hyaC:A, rhaR:A, nac:A, cbl:A, bolA:A were all grown overnite from freezer stocks in 6ml of LB. 3 genes I thought I previously cloned were not in the freezer box (crp, csgD, dctR) [2] One clone, gadX-myc:A didn't grow. This left 13 cultures to be minipreped and sequenced.

All 13 genes were prepped with an eppendorf kit. In addition, 4 genes (bolA, gadX, hyaC, nac) were prepped with the QIAprep spin kit to see how the kits compare (these are shown in italics below).

[2] just discovered [Mon Oct 17 16:05:30 EDT 2005] dctR is actually the same as yhiF so it's not missing

Gene	Amount of Vector (ng/μl)	260/280	260/230	μl for 600 ng
bolA	27.8	1.83	0.57	21.6
bolA	52.6	1.97	2.22	11.4
cbl	33.4	1.96	0.59	18.0
fecI	28.3	1.84	0.68	21.2
flhC	39.1	1.72	0.82	15.3
gadX	35.2	1.75	0.56	17.0
gadX	28.3	2.11	1.98	21.2
hyaC	33.3	1.86	0.58	18.0
hyaC	46.8	1.90	2.04	12.8
nac	30.7	1.80	0.37	19.5
nac	44.5	2.01	2.08	13.5
pdhR	23.1	1.78	0.31	25.8
pdhR-myc	21.0	1.84	0.39	28.6
rhaR	33.9	1.94	0.45	17.7
sfsA	33.3	1.83	0.61	18.0
yheO	34.7	1.84	0.68	17.3
yhiF	28.5	1.75	0.27	21.0

Fourteen samples were sent to Agencourt for sequencing. The bolA, hyaC, and nac plasmids were from the Qiagen kit. bolA was also sent from the Eppy miniprep for comparison (and to check out Agencourt's consistency).

Mon Oct 17 15:51:28 EDT 2005

All samples were digested to reverify insert size (this is mainly because I lost the original image). hyaC was mistakenly taken from the eppy clean up, but everything else is exactly the same as for the sequencing with regards to Eppy and Qiagen. The following protocol was used:

pdhR-myc

Restriction Digest	
vector/DNA	8 μl
EcoRI buffer	2 μl
BSA	2 μl
EcoRI enzyme	0.4 μl
BamHI enzyme	0.4 μl
H_2O	7 μl

all other vectors

Restriction Digest	
vector/DNA	8 μl
EcoRI buffer	2 μl
BSA	0 μl
EcoRI enzyme	0.4 μl
NcoI enzyme	0.4 μl
H_2O	9 μl

Vectors were picked up by Agencourt courier for sequencing on: Mon Oct 17, 2005

Sequences:

```
CLUSTAL W (1.82) multiple sequence alignment
```

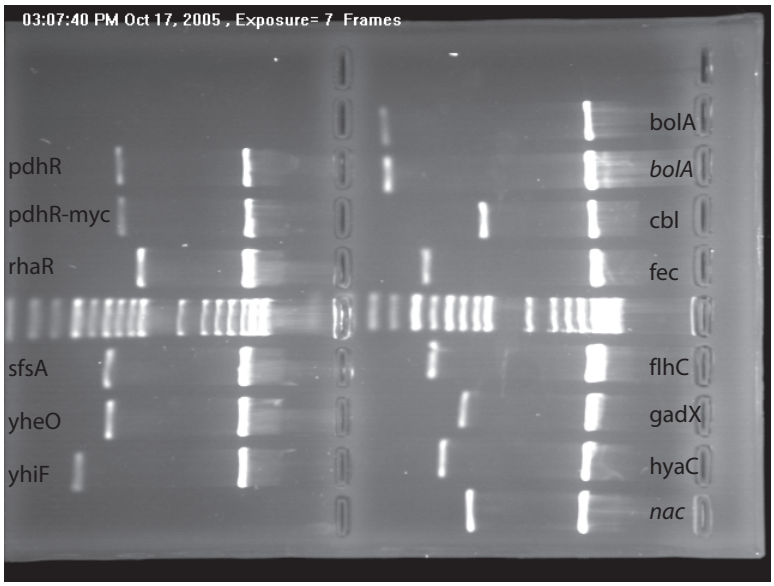


Figure 1.15: 70 ml, 1.5% agarose gel run for 40 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 45 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. 10 μl of Fisher BioReagents exACTGene 1KB Plus DNA Ladder, with bands:weights(ng) of 10000:43, 8000:31, 6000:28, 5000:41, 4000:36, 3000:31, 2500:26, 2000:23, 1500:22, 1000:35, 900:26, 800:52, 700:31, 600:18, 500:35, 400:12, 300:17, 200:12, 100:7 was used. 20ul (all) of each digest was used in 40:1 6x fisher dye. Product sizes for the genes should be: bolA=351bp, cbl=951bp, fec=521bp, flhC=579bp, gadX=825bp, hyaC=708bp, nac=917bp, pdhR=764bp, rhaR=939bp, sfsA=705bp, yheO=735bp, yhiF(dctR)=531bp.

```
bolA_TOPO_pTrcHis.trimmed.seq       GGATCCAACCCTTATGACATCTCAGCGTTGTCGGAGGAGATATTTCATGA 50
bolA_TOPO_pTrcHisB.trimmed.seq      GGATCCA-CCCTTATGA-ATCTCAGCGTTGTCGGAGGAGATATTTCATGA 48
EG10125                             --------------ATGACATCTCAGCGTTGTCGGAGGAGATATTTCATGA 37
                                              **** ********************************

bolA_TOPO_pTrcHis.trimmed.seq       TGATACGTGAGCGGATAGAAGAAAAATTAAGGGCGGCGTTCCAACCCGTA 100
bolA_TOPO_pTrcHisB.trimmed.seq      TGATACGTGAGCGGATAGAAGAAAAATTAAGGGCGGCGTTCCAACCCGTA 98
EG10125                             TGATACGTGAGCGGATAGAAGAAAAATTAAGGGCGGCGTTCCAACCCGTA 87
                                    **************************************************
```

```
bolA_TOPO_pTrcHis.trimmed.seq       TTCCTCGAAGTAGTGGATGAAAGCTATCGTCACAATGTCCCAGCCGGCTT 150
bolA_TOPO_pTrcHisB.trimmed.seq      TTCCTCGAAGTAGTGGATGAAAGCTATCGTCACAATGTCCCAGCCGGCTT 148
EG10125                             TTCCTCGAAGTAGTGGATGAAAGCTATCGTCACAATGTCCCAGCCGGCTC 137
                                    **************************************************

bolA_TOPO_pTrcHis.trimmed.seq       TGAAAGCCATTTTAAAGTTGTGCTGGTCAGCGATCGTTTCACGGGTGAAC 200
bolA_TOPO_pTrcHisB.trimmed.seq      TGAAAGCCATTTTAAAGTTGTGCTGGTCAGCGATCGTTTCACGGGTGAAC 198
EG10125                             TGAAAGCCATTTTAAAGTTGTGCTGGTCAGCGATCGTTTTACGGGTGAAC 187
                                    ************************************* **********

bolA_TOPO_pTrcHis.trimmed.seq       GTTTTCTGAATCGTCATCGAATGATTTACAGTACTTTAGCGGAGGAACTC 250
bolA_TOPO_pTrcHisB.trimmed.seq      GTTTTCTGAATCGTCATCGAATGATTTACAGTACTTTAGCGGAGGAACTC 248
EG10125                             GTTTTCTGAATCGTCATCGAATGATTTACAGTACTTTAGCGGAGGAACTC 237
                                    **************************************************

bolA_TOPO_pTrcHis.trimmed.seq       TCTACTACCGTTCATGCGCTGGCTCTGCATACTTACACTATTAAGGAGTG 300
bolA_TOPO_pTrcHisB.trimmed.seq      TCTACTACCGTTCATGCGCTGGCTCTGCATACTTACACTATTAAGGAGTG 298
EG10125                             TCTACTACCGTTCATGCGCTGGCTCTGCATACTTACACTATTAAGGAGTG 287
                                    **************************************************

bolA_TOPO_pTrcHis.trimmed.seq       GGAAGGGTTGCAGGACACCGTCTTTGCCTCTCCTCCCTGTCGTGGAGCAG 350
bolA_TOPO_pTrcHisB.trimmed.seq      GGAAGGGTTGCAGGACACCGTCTTTGCCTCTCCTCCCTGTCGTGGAGCAG 348
EG10125                             GGAAGGGTTGCAGGACACCGTCTTTGCCTCTCCTCCCTGTCGTGGAGCAG 337
                                    **************************************************

bolA_TOPO_pTrcHis.trimmed.seq       GAAGCATCGCGTAAAAGGGCGAATTCGAAGCTTGGCTGTTTTGGCGGATG 400
bolA_TOPO_pTrcHisB.trimmed.seq      GAAGCATCGCGTAAAAGGGCGAATTCGAAGCTTGGCTGTTTTGGCGGATG 398
EG10125                             GAAGCATCGCGTAA------------------------------------ 351
                                    **************

bolA_TOPO_pTrcHis.trimmed.seq       AGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTC 450
bolA_TOPO_pTrcHisB.trimmed.seq      AGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTC 448
EG10125                             --------------------------------------------------

bolA_TOPO_pTrcHis.trimmed.seq       TGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGAC 500
bolA_TOPO_pTrcHisB.trimmed.seq      TGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGAC 498
EG10125                             --------------------------------------------------

bolA_TOPO_pTrcHis.trimmed.seq       CCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGG 550
bolA_TOPO_pTrcHisB.trimmed.seq      CCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGG 548
EG10125                             --------------------------------------------------

bolA_TOPO_pTrcHis.trimmed.seq       GTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAG 600
bolA_TOPO_pTrcHisB.trimmed.seq      GTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAG 598
EG10125                             --------------------------------------------------

bolA_TOPO_pTrcHis.trimmed.seq       GCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAA 650
bolA_TOPO_pTrcHisB.trimmed.seq      GCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAA 648
EG10125                             --------------------------------------------------

bolA_TOPO_pTrcHis.trimmed.seq       CGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGA 700
bolA_TOPO_pTrcHisB.trimmed.seq      CGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGA 698
EG10125                             --------------------------------------------------

bolA_TOPO_pTrcHis.trimmed.seq       AGCAACGGCCCGGAAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAG 750
bolA_TOPO_pTrcHisB.trimmed.seq      AGCAACGGCCCGGAGGGGTGGCGGGC------------------------- 724
EG10125                             --------------------------------------------------

bolA_TOPO_pTrcHis.trimmed.seq       GCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTC 800
bolA_TOPO_pTrcHisB.trimmed.seq      --------------------------------------------------
EG10125                             --------------------------------------------------

bolA_TOPO_pTrcHis.trimmed.seq       TACAAACTCTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCT 850
bolA_TOPO_pTrcHisB.trimmed.seq      --------------------------------------------------
EG10125                             --------------------------------------------------

bolA_TOPO_pTrcHis.trimmed.seq       CATGAGACAATAACCCTGATAAATGCTTTCATAATATTGA 890
bolA_TOPO_pTrcHisB.trimmed.seq      ----------------------------------------
EG10125                             ----------------------------------------


CLUSTAL W (1.82) multiple sequence alignment


cbl_TOPO_pTrcHis.trimmed.seq      GGATCCAACCCTTGTGAATTTCCA-CAACTAAAGATAATCCGCGAGGCTG 49
G7071                             ------------GTGAATTTCCAACAACTAAAGATAATCCGCGAGGCTG 37
                                            *********** ************************

cbl_TOPO_pTrcHis.trimmed.seq      CACGTCAGGATTACAACCTGACAGAGGTTGCGAATATGCTTTTTACCTCA 99
G7071                             CACGTCAGGATTACAACCTGACAGAGGTTGCGAATATGCTTTTTACCTCA 87
                                  **************************************************

cbl_TOPO_pTrcHis.trimmed.seq      CAGTCAGGCGTCAGCCGTCATATTCGGGAACTGGAGGATGAACTTGGCAT 149
```

```
G7071                           CAGTCAGGCGTCAGCCGTCATATTCGGGAACTGGAGGATGAACTTGGCAT 137
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    CGAAATATTTGTTCGACGAGGTAAGCGACTGCTGGGCATGACTGAACCGG 199
G7071                           CGAAATATTTGTTCGACGAGGTAAGCGACTGCTGGGCATGACTGAACCGG 187
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    GCAAAGCATTACTGGTCATTGCAGAACGTATTCTGAATGAAGCCAGTAAT 249
G7071                           GCAAAGCATTACTGGTCATTGCAGAACGTATTCTGAATGAAGCCAGTAAT 237
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    GTTCGTCGGCTTGCAGACCTGTTTACCAACGATACGTCTGGCGTTCTCAC 299
G7071                           GTTCGTCGGCTTGCAGACCTGTTTACCAACGATACGTCTGGCGTTCTCAC 287
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    TATTGCAACGACGCATACTCAGGCACGTTATAGCTTGCCAGAGGTCATTA 349
G7071                           TATTGCAACGACGCATACTCAGGCACGTTATAGCTTGCCAGAGGTCATTA 337
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    AAGCTTTTCGCGAACTTTTCCCGGAGGTTCGGCTCGAGCTAATCCAGGGG 399
G7071                           AAGCTTTTCGCGAACTTTTCCCGGAGGTTCGGCTCGAGCTAATCCAGGGG 387
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    ACGCCACAGGAAATTGCGACATTGTTGCAAAATGGCGAAGCTGATATTGG 449
G7071                           ACGCCACAGGAAATTGCGACATTGTTGCAAAATGGCGAAGCTGATATTGG 437
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    TATCGCCAGCGAGCGTTTGAGTAATGACCCGCAGCTCGTCGCCTTCCCGT 499
G7071                           TATCGCCAGCGAGCGTTTGAGTAATGACCCGCAGCTCGTCGCCTTCCCGT 487
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    GGTTTCGTTGGCACCATAGTTTGCTTGTTCCACACGATCATCCCTTGACG 549
G7071                           GGTTTCGTTGGCACCATAGTTTGCTTGTTCCACACGATCATCCCTTGACG 537
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    CAAATTTCACCATTGACGCTGGAATCAATAGCGAAGTGGCCGTTAATCAC 599
G7071                           CAAATTTCACCATTGACGCTGGAATCAATAGCGAAGTGGCCGTTAATCAC 587
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    TTACCGACAGGGGATTACGGGGCGCTCACGTATTGATGACGCATTTGCCC 649
G7071                           TTACCGACAGGGGATTACGGGGCGCTCACGTATTGATGACGCATTTGCCC 637
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    GCAAAGGTTTGCTGGCAGATATTGTATTAAGTGCGCAGGATTCTGATGTC 699
G7071                           GCAAAGGTTTGCTGGCAGATATTGTATTAAGTGCGCAGGATTCTGATGTC 687
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    ATTAAAACCTATGTTGCTCTTGGGCTTGGGATCGGATTAGTTGCCGAGCA 749
G7071                           ATTAAAACCTATGTTGCTCTTGGGCTTGGGATCGGATTAGTTGCCGAGCA 737
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    ATCCAGTGGCGAACAAGAGGAAGAGAATTTAATCCGCCTGGATACGCGGC 799
G7071                           ATCCAGTGGCGAACAAGAGGAAGAGAATTTAATCCGCCTGGATACGCGGC 787
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    ATCTTTTTGATGCTAATACTGTCTGGTTGGGACTGAAGCGAGGACAACTT 849
G7071                           ATCTTTTTGATGCTAATACTGTCTGGTTGGGACTGAAGCGAGGACAACTT 837
                                **************************************************

cbl_TOPO_pTrcHis.trimmed.seq    CAGCGTAACTATGTCTGGCGCTTTCTGGAACT------------------ 881
G7071                           CAGCGTAACTATGTCTGGCGCTTTCTGGAACTTTGTAATGCAGGACTGTC 887
                                **********************************

cbl_TOPO_pTrcHis.trimmed.seq    --------------------------------------------------
G7071                           AGTTGAGGATATCAAGCGGCAGGTGATGGAAAGCAGTGAAGAGGAAATTG 937


cbl_TOPO_pTrcHis.trimmed.seq    --------------
G7071                           ATTATCAGATATAG 951


CLUSTAL W (1.82) multiple sequence alignment

REVERSED!!!!!


flhC_TOPO_pTrcHis.trimmed.seq      CAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCAGGAGAGCGTTC 50
EG10319                            --------------------------------------------------


flhC_TOPO_pTrcHis.trimmed.seq      ACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTTCGACTGAGCCTT 100
EG10319                            --------------------------------------------------


flhC_TOPO_pTrcHis.trimmed.seq      TCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCATGGGGAGACCCC 150
EG10319                            --------------------------------------------------


flhC_TOPO_pTrcHis.trimmed.seq      ACACTACCATCGGCGCTACGGCGTTTCACTTCTGAGTTCGGCATGGGGTC 200
EG10319                            --------------------------------------------------
```

```
flhC_TOPO_pTrcHis.trimmed.seq      AGGTGGGACCACCGCGCTACTGCCGCCAGGCAAATTCTGTTTTATCAGAC 250
EG10319                            --------------------------------------------------

flhC_TOPO_pTrcHis.trimmed.seq      CGCTTCTGCGTTCTGATTTAATCTGTATCAGGCTGAAAATCTTCTCTCAT 300
EG10319                            --------------------------------------------------

flhC_TOPO_pTrcHis.trimmed.seq      CCGCCAAAACAGCCAAGCTTCGAATTCGCCCTTATGAGTGAAAAAAGCAT 350
EG10319                            ----------------------------------ATGAGTGAAAAAAGCAT 17
                                                                      ******************

flhC_TOPO_pTrcHis.trimmed.seq      TGTTCAGGAAGCGCGGGATATTCAGCTGGCAATGGAATTGATCACCCTGG 400
EG10319                            TGTTCAGGAAGCGCGGGATATTCAGCTGGCAATGGAATTGATCACCCTGG 67
                                   **************************************************

flhC_TOPO_pTrcHis.trimmed.seq      GCGCTCGTTTGCAGATGCTGGAAAGCGAAACACAGTTAAGTCGCGGACGC 450
EG10319                            GCGCTCGTTTGCAGATGCTGGAAAGCGAAACACAGTTAAGTCGCGGACGC 117
                                   **************************************************

flhC_TOPO_pTrcHis.trimmed.seq      CTGATAAAACTTTATAAAGAACTGCGCGGAAGCCCACCGCCGAAAGGCAT 500
EG10319                            CTGATAAAACTTTATAAAGAACTGCGCGGAAGCCCACCGCCGAAAGGCAT 167
                                   **************************************************

flhC_TOPO_pTrcHis.trimmed.seq      GCTGCCATTCTCAACCGACTGGTTTATGACCTGGGAACAAAACGTTCATG 550
EG10319                            GCTGCCATTCTCAACCGACTGGTTTATGACCTGGGAACAAAACGTTCATG 217
                                   **************************************************

flhC_TOPO_pTrcHis.trimmed.seq      CTTCGATGTTCTGTAATGCATGGCAGTTTTTACTGAAAACCGGTTTGTGT 600
EG10319                            CTTCGATGTTCTGTAATGCATGGCAGTTTTTACTGAAAACCGGTTTGTGT 267
                                   **************************************************

flhC_TOPO_pTrcHis.trimmed.seq      AATGGCGTCGATGCGGTGATCAAAGCCTACCGTTTATACCTTGAACAGTG 650
EG10319                            AATGGCGTCGATGCGGTGATCAAAGCCTACCGTTTATACCTTGAACAGTG 317
                                   **************************************************

flhC_TOPO_pTrcHis.trimmed.seq      CCCACAAGCAGAAGAAGGACCACTGCTGGCATTAACCCGTGCCTGGACAT 700
EG10319                            CCCACAAGCAGAAGAAGGACCACTGCTGGCATTAACCCGTGCCTGGACAT 367
                                   **************************************************

flhC_TOPO_pTrcHis.trimmed.seq      TGGTGCGGTTTGTTGAAAGTGGATTACTGCAACTTTCCAGCTGCAACTGC 750
EG10319                            TGGTGCGGTTTGTTGAAAGTGGATTACTGCAACTTTCCAGCTGCAACTGC 417
                                   **************************************************

flhC_TOPO_pTrcHis.trimmed.seq      TGCGGCGGCAATTTTATTACCCACGCTCACCAGCCTGTTGGCAGCTTTGC 800
EG10319                            TGCGGCGGCAATTTTATTACCCACGCTCACCAGCCTGTTGGCAGCTTTGC 467
                                   **************************************************

flhC_TOPO_pTrcHis.trimmed.seq      CTGCAGCTTATGTCAACCGCCATCCCGGGCAGTAAAAAGACGTAAACTTT 850
EG10319                            CTGCAGCTTATGTCAACCGCCATCCCGGGCAGTAAAAAGACGTAAACTTT 517
                                   **************************************************

flhC_TOPO_pTrcHis.trimmed.seq      CCCAGAATCCTGCCGATATTATCCCACAACTGCAGGATGAACAGAGAGTA 900
EG10319                            CCCAGAATCCTGCCGATATTATCCCACAACTGCTGGATGAACAGAGAGTA 567
                                   ********************************** ****************

flhC_TOPO_pTrcHis.trimmed.seq      CAGGCTGTTAAAAGGG 916
EG10319                            CAGGCTGTTTAA---- 579
                                   ********* **


CLUSTAL W (1.82) multiple sequence alignment


gadX_TOPO_pTrcHis.trimmed.seq      GGATCCAACCCTTATGCAATCACTACATGGGAATTGTCTAATTGCGTATG 50
EG12243                            ------------ATGCAATCACTACATGGGAATTGTCTAATTGCGTATG 37
                                               ************************************

gadX_TOPO_pTrcHis.trimmed.seq      CAAGACATAAATATATTCTCACCATGGTTAATGGTGAATATCGCTATTTT 100
EG12243                            CAAGACATAAATATATTCTCACCATGGTTAATGGTGAATATCGCTATTTT 87
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      AATGGCGGTGACCTGGTTTTTGCGGATACAAGCCAAATTCGAGTAGATAA 150
EG12243                            AATGGCGGTGACCTGGTTTTTGCGGATGCAAGCCAAATTCGAGTAGATAA 137
                                   ************************** **********************

gadX_TOPO_pTrcHis.trimmed.seq      GTGTGTTGAAAATTTTGTATTCGTGTCAAGGGACACGCTTTCATTATTTC 200
EG12243                            GTGTGTTGAAAATTTTGTATTCGTGTCAAGGGACACGCTTTCATTATTTC 187
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      TCCCGATGCTCAAGGAGGAGGCATTAAATCTTCATGCACATAAAAAAGTT 250
EG12243                            TCCCGATGCTCAAGGAGGAGGCATTAAATCTTCATGCACATAAAAAAGTT 237
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      TCTTCATTACTCGTTCATCACTGTAGTAGAGATATTCCTGTTTTTCAGGA 300
EG12243                            TCTTCATTACTCGTTCATCACTGTAGTAGAGATATTCCTGTTTTTCAGGA 287
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      AGTTGCGCAACTATCGCAGAATAAGAATCTTCGCTATGCAGAAATGCTAC 350
EG12243                            AGTTGCGCAACTATCGCAGAATAAGAATCTTCGCTATGCAGAAATGCTAC 337
                                   **************************************************
```

```
gadX_TOPO_pTrcHis.trimmed.seq      GTAAAAGAGCATTAATCTTTGCGTTGTTATCTGTTTTTCTTGAGGATGAG 400
EG12243                            GTAAAAGAGCATTAATCTTTGCGTTGTTATCTGTTTTTCTTGAGGATGAG 387
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      CACTTTATACCGCTGCTTCTGAACGTTTTACAACCGAACATGCGAACACG 450
EG12243                            CACTTTATACCGCTGCTTCTGAACGTTTTACAACCGAACATGCGAACACG 437
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      AGTTTGTACGGTTATCAATAATAATATCGCCCATGAGTGGACACTAGCCC 500
EG12243                            AGTTTGTACGGTTATCAATAATAATATCGCCCATGAGTGGACACTAGCCC 487
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      GAATCGCCAGCGAGCTGTTGATGAGTCCAAGTCTGTTAAAGAAAAAATTG 550
EG12243                            GAATCGCCAGCGAGCTGTTGATGAGTCCAAGTCTGTTAAAGAAAAAATTG 537
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      CGCGAAGAAGAGACATCATATTCACAGTTGCTTACTGAGTGTAGAATGCA 600
EG12243                            CGCGAAGAAGAGACATCATATTCACAGTTGCTTACTGAGTGTAGAATGCA 587
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      ACGTGCTTTGCAACTTATTGTTATACATGGTTTTTCAATTAAGCGAGTTG 650
EG12243                            ACGTGCTTTGCAACTTATTGTTATACATGGTTTTTCAATTAAGCGAGTTG 637
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      CAGTATCCTGTGGATATCACAGCGTGTCGTATTTCATTTACGTCTTTCGA 700
EG12243                            CAGTATCCTGTGGATATCACAGCGTGTCGTATTTCATTTACGTCTTTCGA 687
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      AATTATTATGGGATGACGCCCACAGAGTATCAGGAGCGATCGGCGCAGAG 750
EG12243                            AATTATTATGGGATGACGCCCACAGAGTATCAGGAGCGATCGGCGCAGAG 737
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      ATTGTCGAACCGTGACTCGGCGGCAAGTATTGTTGCGCAAGGGAATTTTT 800
EG12243                            ATTGTCGAACCGTGACTCGGCGGCAAGTATTGTTGCGCAAGGGAATTTTT 787
                                   **************************************************

gadX_TOPO_pTrcHis.trimmed.seq      ACGGCACTGACCGTTCTGCGGAAGGAATAAGATTATAGAAGGGCGAATTC 850
EG12243                            ACGGCACTGACCGTTCTGCGGAAGGAATAAGATTATAG------------ 825
                                   ************************************

gadX_TOPO_pTrcHis.trimmed.seq      GAAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGA 900
EG12243                            -------------------------------------------------
```

```
CLUSTAL W (1.82) multiple sequence alignment

MISSING LAST 2 CODONS (not human error)
```

```
hyaC_TOPO_pTrcHisB.trimmed.seq      CCCTTATGCAACAGAAAAGCGACAACGTTGTCAGCCACTATGTCTTTGAA 50
EG10470                             -----ATGCAACAGAAAAGCGACAACGTTGTCAGCCACTATGTCTTTGAA 45
                                         *********************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      GCGCCAGTGCGCATCTGGCACTGGTTGACGGTGTTATGCATGGCGGTGTT 100
EG10470                             GCGCCAGTGCGCATCTGGCACTGGTTGACGGTGTTATGCATGGCGGTGTT 95
                                    **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      GATGGTCACCGGATACTTTATCGGCAAGCCGCTACCTTCCGTCAGCGGCG 150
EG10470                             GATGGTCACCGGATACTTTATCGGCAAGCCGCTACCTTCCGTCAGCGGCG 145
                                    **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      AGGCGACGTATCTGTTCTATATGGGCTACATCAGGTTAATTCACTTCAGC 200
EG10470                             AGGCGACGTATCTGTTCTATATGGGCTACATCAGGTTAATTCACTTCAGC 195
                                    **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      GCCGGGATGGTTTTTACCGTGGTTTTGCTGATGCGGATCTACTGGGCTTT 250
EG10470                             GCCGGGATGGTTTTTACCGTGGTTTTGCTGATGCGGATCTACTGGGCTTT 245
                                    **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      TGTTGGCAATCGATACTCCCGCGAGCTGTTTATCGTGCCGGTATGGCGTA 300
EG10470                             TGTTGGCAATCGATACTCCCGCGAGCTGTTTATCGTGCCGGTATGGCGTA 295
                                    **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      AAAGCTGGTGGCAGGGCGTGTGGTATGAAATCCGCTGGTATCTGTTTCTG 350
EG10470                             AAAGCTGGTGGCAGGGCGTGTGGTATGAAATCCGCTGGTATCTGTTTCTG 345
                                    **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      GCAAAACGTCCGAGTGCCGATATAGGCCATAATCCCATCGCCCAGGCGGC 400
EG10470                             GCAAAACGTCCGAGTGCCGATATAGGCCATAATCCCATCGCCCAGGCGGC 395
                                    **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      GATGTTCGGCTATTTCCTGATGTCGGTCTTTATGATCATCACTGGTTTTG 450
EG10470                             GATGTTCGGCTATTTCCTGATGTCGGTCTTTATGATCATCACTGGTTTTG 445
                                    **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      CGCTGTACAGCGAACACAGCCAGTACGCTATTTTTGCGCCGTTCCGTTAT 500
EG10470                             CGCTGTACAGCGAACACAGCCAGTACGCTATTTTTGCGCCGTTCCGTTAT 495
                                    **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      GTGGTGGAATTTTTCTACTGGACGGGTGGCAACTCAATGGACATTCACAG 550
EG10470                             GTGGTGGAATTTTTCTACTGGACGGGTGGCAACTCAATGGACATTCACAG 545
                                    **************************************************
```

```
hyaC_TOPO_pTrcHisB.trimmed.seq      CTGGCATCGGCTGGGGATGTGGCTGATTGGCGCGTTTGTGATCGGTCATG 600
EG10470                            CTGGCATCGGCTGGGGATGTGGCTGATTGGCGCGTTTGTGATCGGTCATG 595
                                   **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      TCTACATGGCGCTGCGTGAAGACATCATGTCCGACGACACGGTGATCTCC 650
EG10470                            TCTACATGGCGCTGCGTGAAGACATCATGTCCGACGACACGGTGATCTCC 645
                                   **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      ACCATGGTCAACGGCTACCGTAGCCACAAATTTGGCAAAATAAGTAACAA 700
EG10470                            ACCATGGTCAACGGCTACCGTAGCCACAAATTTGGCAAAATAAGTAACAA 695
                                   **************************************************

hyaC_TOPO_pTrcHisB.trimmed.seq      GGAGCGTAAGGGCGAATTCGAAGCTTGGCTGTTTTGGCGGATGAGAGAAG 750
EG10470                            GGAGCGTTCATGA---------------------------------------- 708
                                   *******    *

hyaC_TOPO_pTrcHisB.trimmed.seq      ATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGANTAA 800
EG10470                            --------------------------------------------------


hyaC_TOPO_pTrcHisB.trimmed.seq      ACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGC 850
EG10470                            --------------------------------------------------


hyaC_TOPO_pTrcHisB.trimmed.seq      CGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCC 900
EG10470                            --------------------------------------------------


hyaC_TOPO_pTrcHisB.trimmed.seq      CATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAACGAAAGGCTCAGTC 950
EG10470                            --------------------------------------------------


hyaC_TOPO_pTrcHisB.trimmed.seq      GAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCG 987
EG10470                            -------------------------------------


CLUSTAL W (1.82) multiple sequence alignment


pdhR_TOPO_pTrcHis.trimmed.seq      --------------------------------------------------
EG11088                            --------------------------------------------------
his2                               TTATCGATTAACTTTATTATAAAAATTAAAGAGGTATATATTAATGTATC 50


pdhR_TOPO_pTrcHis.trimmed.seq      -----------------GGATCCA---CCCTTATGGCCTACAGCAAAATC 30
EG11088                            --------------------------------ATGGCCTACAGCAAAATC 18
his2                               GATTAAATAAGGAGGAATAAACCATGGCCCTTATGGCCTACAGCAAAATC 100
                                                                   ******************

pdhR_TOPO_pTrcHis.trimmed.seq      CGCCAACCAATACTCTCCGATGTGATTGAGCAGCAACTGGAGTTTTTGAT 80
EG11088                            CGCCAACCAAAACTCTCCGATGTGATTGAGCAGCAACTGGAGTTTTTGAT 68
his2                               CGCCAACCAAAACTCTCCGATGTGATGNAGCAGCAACTGGAGTTTTTGAT 150
                                   ********** ***************  **********************

pdhR_TOPO_pTrcHis.trimmed.seq      CCTCGAAGGCACTCTCCGCCCGGGCGAAAAACTCCCACCGGAACGCGAAC 130
EG11088                            CCTCGAAGGCACTCTCCGCCCGGGCGAAAAACTCCCACCGGAACGCGAAC 118
his2                               CCTCGAAGGCACTCTCCGCCCGGGCGAAAAACTCCCACCGGAACGCGAAC 200
                                   **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq      TGGCAAAACAGTTTGACGTCTCCCGTCCCTCCTTGCGTGAGGCGATTCAA 180
EG11088                            TGGCAAAACAGTTTGACGTCTCCCGTCCCTCCTTGCGTGAGGCGATTCAA 168
his2                               TGGCAAAACAGTTTGACGTCTCCCGTCCCTCCTTGCGTGAGGCGATTCAA 250
                                   **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq      CGTCTCGAAGCGAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTT 230
EG11088                            CGTCTCGAAGCGAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTT 218
his2                               CGTCTCGAAGCGAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTT 300
                                   **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq      TGTCCAGAGCAGCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGC 280
EG11088                            TGTCCAGAGCAGCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGC 268
his2                               TGTCCAGAGCAGCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGC 350
                                   **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq      TCTCCGACCATCCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCC 330
EG11088                            TCTCCGACCATCCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCC 318
his2                               TCTCCGACCATCCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCC 400
                                   **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq      CTGGAAGGTATCGCCGCTTATTACGCCGCGCTGCGTAGTACCGATGAAGA 380
EG11088                            CTGGAAGGTATCGCCGCTTATTACGCCGCGCTGCGTAGTACCGATGAAGA 368
his2                               CTGGAAGGTATCGCCGCTTATTACGCCGCGCTGCGTAGTACCGATGAAGA 450
                                   **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq      CAAGGAACGCATCCGTGAACTCCACCACGCCATAGAGCTGGCGCAGCAGT 430
EG11088                            CAAGGAACGCATCCGTGAACTCCACCACGCCATAGAGCTGGCGCAGCAGT 418
his2                               CAAGGAACGCATCCGTGAACTCCACCACGCCATAGAGCTGGCGCAGCAGT 500
                                   **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq      CTGGCGATCTGGACGCGGAATCAAACGCCGTACTCCAGTATCAGATTGCC 480
EG11088                            CTGGCGATCTGGACGCGGAATCAAACGCCGTACTCCAGTATCAGATTGCC 468
```

```
his2                              CTGGCGATCTGGACGCGGAATCAAACGCCGTACTCCAGTATCAGATTGCC 550
                                  **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq     GTCACCGAAGCGGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTAT 530
EG11088                           GTCACCGAAGCGGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTAT 518
his2                              GTCACCGAAGCGGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTAT 600
                                  **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq     GGAGCCGATGTTGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATT 580
EG11088                           GGAGCCGATGTTGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATT 568
his2                              GGAGCCGATGTTGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATT 650
                                  **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq     CGCGTCGCGAGATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTT 630
EG11088                           CGCGTCGCGAGATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTT 618
his2                              CGCGTCGCGAGATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTT 700
                                  **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq     GAAGCGATTATGGCCGGTAAGCCGGAAGAAGCGCGCGAAGCATCGCATCG 680
EG11088                           GAAGCGATTATGGCCGGTAAGCCGGAAGAAGCGCGCGAAGCATCGCATCG 668
his2                              GAAGCGATTATGGCCGGTAAGCCGGAAGAAGCGCGCGAAGCATCGCATCG 750
                                  **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq     CCATCTGGCCTTTATCGAAGAAATTTTGCTCGACAGAAGTCGTGAAGAGA 730
EG11088                           CCATCTGGCCTTTATCGAAGAAATTTTGCTCGACAGAAGTCGTGAAGAGA 718
his2                              CCATCTGGCCTTTATCGAAGAAATTTTGCTCGACAGAAGTCGTGAAGAGA 800
                                  **************************************************

pdhR_TOPO_pTrcHis.trimmed.seq     GCCGCCGTGAGCGTTCTCTGCGTCGTCTGGAGCAACGAAAGAATTAGAAG 780
EG11088                           GCCGCCGTGAGCGTTCTCTGCGTCGTCTGGAGCAACGAAAGAATTAG--- 765
his2                              GCCGCCGTGAGCGTTCTCTGCGTCGTCTGGAGCAACGAAAGAAT---AAG 847
                                  ******************************************

pdhR_TOPO_pTrcHis.trimmed.seq     GGCGAATTCGAAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCC 830
EG11088                           -------------------------------------------------- 
his2                              GGCGAATTCGAAGCTTACGTA-------GAACAAAAACTCATCTCAGAAG 890


pdhR_TOPO_pTrcHis.trimmed.seq     TGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTG 880
EG11088                           -------------------------------------------------- 
his2                              AGGATCTGAATAGCGCCGTCGACCATCATCATC-ATCATCATTGAGTTAA 939


pdhR_TOPO_pTrcHis.trimmed.seq     CCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACT 926
EG11088                           ----------------------------------------------- 
his2                              ACGGTCTCCAG---------------------------------- 950


CLUSTAL W (1.82) multiple sequence alignment

REVERSED!!!!!


rhaR_TOPO_pTrcHis.trimmed.seq     ------------------------------CAACGTATTTGTACGCCATAT 21
EG10842                           ATGGCTTTCTGCAATAACGCGAATCTTCTCAACGTATTTGTACGCCATAT 50
                                                               ********************

rhaR_TOPO_pTrcHis.trimmed.seq     TGCGAATAATCAACTTCGTTCTCTGGCCGAGGTAGCCACGGTGGCGCATC 71
EG10842                           TGCGAATAATCAACTTCGTTCTCTGGCCGAGGTAGCCACGGTGGCGCATC 100
                                  **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq     AGTTAAAACTTCTCAAAGATGATTTTTTTGCCAGCGACCAGCAGGCAGTC 121
EG10842                           AGTTAAAACTTCTCAAAGATGATTTTTTTGCCAGCGACCAGCAGGCAGTC 150
                                  **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq     GCTGTGGCTGACCGTTATCCGCAAGATGTCTTTGCTGAACATACACATGA 171
EG10842                           GCTGTGGCTGACCGTTATCCGCAAGATGTCTTTGCTGAACATACACATGA 200
                                  **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq     TTTTTGTGAGCTGGTGATTGTCTGGCGCGGTAATGGCCTGCATGTACTCA 221
EG10842                           TTTTTGTGAGCTGGTGATTGTCTGGCGCGGTAATGGCCTGCATGTACTCA 250
                                  **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq     ACGATCGCCCTTATCGCATTACCCGTGGCGATCTCTTTTACATTCATGCT 271
EG10842                           ACGATCGCCCTTATCGCATTACCCGTGGCGATCTCTTTTACATTCATGCT 300
                                  **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq     GACGATAAACACTCCTACGCTTCCGTTAACGATCTGGTTTTGCAGAATAT 321
EG10842                           GACGATAAACACTCCTACGCTTCCGTTAACGATCTGGTTTTGCAGAATAT 350
                                  **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq     TATTTATTGCCCGGAGCGTCTGAAGCTGAATCTTGACTGGCAGGGGGCGA 371
EG10842                           TATTTATTGCCCGGAGCGTCTGAAGCTGAATCTTGACTGGCAGGGGGCGA 400
                                  **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq     TTCCGGGATTTAACGCCAGCGCAGGGCAACCACACTGGCGCTTAGGTAGC 421
EG10842                           TTCCGGGATTTAACGCCAGCGCAGGGCAACCACACTGGCGCTTAGGTAGC 450
                                  **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq     ATGGGGATGGCGCAGGCGCGGCAGGTTATCGGTCAGCTTGAGCATGAAAG 471
EG10842                           ATGGGGATGGCGCAGGCGCGGCAGGTTATCGGTCAGCTTGAGCATGAAAG 500
                                  **************************************************
```

```
rhaR_TOPO_pTrcHis.trimmed.seq      TAGTCAGCATGTGCCGTTTGCTAACGAAATGGCTGAGTTGCTGTTCGGGC 521
EG10842                            TAGTCAGCATGTGCCGTTTGCTAACGAAATGGCTGAGTTGCTGTTCGGGC 550
                                   **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq      AGTTGGTGATGTTGCTGAATCGCCATCGTTACACCAGTGATTCGTTGCCG 571
EG10842                            AGTTGGTGATGTTGCTGAATCGCCATCGTTACACCAGTGATTCGTTGCCG 600
                                   **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq      CCAACATCCAGCGAAACGTTGCTGGATAAGCTGATTACCCGGCTGGCGGC 621
EG10842                            CCAACATCCAGCGAAACGTTGCTGGATAAGCTGATTACCCGGCTGGCGGC 650
                                   **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq      TAGCCTGAAAAGTCCCTTTGCGCTGGATAAATTTTGTGATGAGGCATCGT 671
EG10842                            TAGCCTGAAAAGTCCCTTTGCGCTGGATAAATTTTGTGATGAGGCATCGT 700
                                   **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq      GCAGTGAGCGCGTTTTGCGTCAGCAATTTCGCCAGCAGACTGGAATGACC 721
EG10842                            GCAGTGAGCGCGTTTTGCGTCAGCAATTTCGCCAGCAGACTGGAATGACC 750
                                   **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq      ATCAATCAATATCTGCGACAGGTCAGAGTGTGTCATGCGCAATATCTTCT 771
EG10842                            ATCAATCAATATCTGCGACAGGTCAGAGTGTGTCATGCGCAATATCTTCT 800
                                   **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq      CCAGCATAGCCGCCTGTTAATCAGTGATATTTCGACCGAATGTGGCTTTG 821
EG10842                            CCAGCATAGCCGCCTGTTAATCAGTGATATTTCGACCGAATGTGGCTTTG 850
                                   **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq      AAGATAGTAACTATTTTTCGGTGGTGTTTACCCGGGAAACCGGGATGACG 871
EG10842                            AAGATAGTAACTATTTTTCGGTGGTGTTTACCCGGGAAACCGGGATGACG 900
                                   **************************************************

rhaR_TOPO_pTrcHis.trimmed.seq      CCCAGCCAGTGGCGTCATCTCAATTCGCAGAAAGATAAAAGGG 914
EG10842                            CCCAGCCAGTGGCGTCATCTCAATTCGCAGAAAGATTAA---- 939
                                   ********************************** **


CLUSTAL W (1.82) multiple sequence alignment

REVERSED!!!


sfsA_TOPO_pTrcHis.trimmed.seq      GAGCCTTTCGTTTTATTGATGCTGGGCAGTCCCCTACTCTCGCATGGGGA 50
EG10949                            --------------------------------------------------


sfsA_TOPO_pTrcHis.trimmed.seq      GACCCCACACTACCATCGGCGCTACGGCGTTTCACTTCTGAGTTCGGCAT 100
EG10949                            --------------------------------------------------


sfsA_TOPO_pTrcHis.trimmed.seq      GGGGTCAGGTGGGACCACCGCGCTACTGCCGCCAGGCAAATTCTGTTTTA 150
EG10949                            --------------------------------------------------


sfsA_TOPO_pTrcHis.trimmed.seq      TCAGACCGCTTCTGCGTTCTGATTTAATCTGTATCAGGCTGAAAATCTTC 200
EG10949                            --------------------------------------------------


sfsA_TOPO_pTrcHis.trimmed.seq      TCTCATCCGCCAAAACAGCCAAGCTTCGAATTCGCCCTTATGGAATTTTC 250
EG10949                            ----------------------------------------ATGGAATTTTC 11
                                                                           ***********

sfsA_TOPO_pTrcHis.trimmed.seq      TCCCCCTCTACAGCGCGCGACGCTAATTCAGCGTTACAAACGTTTTTTAG 300
EG10949                            TCCCCCTCTACAGCGCGCGACGCTAATTCAGCGTTACAAACGTTTTTTAG 61
                                   **************************************************

sfsA_TOPO_pTrcHis.trimmed.seq      CCGATGTGATCACACCCGATGGTCGCGAATTAACGCTACACTGCCCGAAT 350
EG10949                            CCGATGTGATCACACCCGATGGTCGCGAATTAACGCTACACTGCCCGAAT 111
                                   **************************************************

sfsA_TOPO_pTrcHis.trimmed.seq      ACGGGTGCGATGACCGGTTGCGCAACGCCTGGCGATACCGTCTGGTATTC 400
EG10949                            ACGGGTGCGATGACCGGTTGTGCAACGCCTGGCGATACCGTCTGGTATTC 161
                                   ****************** ****************************

sfsA_TOPO_pTrcHis.trimmed.seq      GACTTCAGACAACACCAAACGGAAATACCCACACACCTGGGAATTAACTC 450
EG10949                            GACTTCAGACAACACCAAACGGAAATACCCACACACCTGGGAATTAACTC 211
                                   **************************************************

sfsA_TOPO_pTrcHis.trimmed.seq      AAAGCCAGAGCGGCGCATTTATTTGCGTCAACACGCTTTGGGCTAACAGG 500
EG10949                            AAAGCCAGAGCGGCGCATTTATTTGCGTCAACACGCTTTGGGCTAACAGG 261
                                   **************************************************

sfsA_TOPO_pTrcHis.trimmed.seq      TTGACGAAAGAGGCTATCCTTAATGAATCAATTTCAGAACTGTCAGGCTA 550
EG10949                            TTGACGAAAGAGGCTATCCTTAATGAATCAATTTCAGAACTGTCAGGCTA 311
                                   **************************************************

sfsA_TOPO_pTrcHis.trimmed.seq      TAGCTCGCTGAAAAGCGAAGTAAAATACGGCGCAGAACGCAGCCGTATTG 600
EG10949                            TAGCTCGCTGAAAAGCGAAGTAAAATACGGCGCAGAACGCAGCCGTATTG 361
                                   **************************************************

sfsA_TOPO_pTrcHis.trimmed.seq      ACTTTATGTTGCAGGCGGATTCGCATCCAGACTGCTATATTGAAGTGAAA 650
```

```
EG10949                         ACTTTATGTTGCAGGCGGATTCGCGTCCAGACTGCTATATTGAAGTGAAA 411
                                ********************** ************************

sfsA_TOPO_pTrcHis.trimmed.seq   TCGGTTACGTTAGCGGAGAACGAACAGGGATATTTTCCCGATGCGGTCAC 700
EG10949                         TCGGTTACGTTAGCGGAGAACGAACAGGGATATTTTCCCGATGCGGTCAC 461
                                **************************************************

sfsA_TOPO_pTrcHis.trimmed.seq   TGAACGAGGTCAGAAACACCTTCGGGAGTTGATGAGCGTAGCGGCTGAAG 750
EG10949                         TGAACGAGGTCAGAAACACCTTCGGGAGTTGATGAGCGTAGCGGCTGAAG 511
                                **************************************************

sfsA_TOPO_pTrcHis.trimmed.seq   GCCAGCGTGCGGTTATCTTTTTCGCCGTGCTGCATTCAGCCATTACACGG 800
EG10949                         GCCAGCGTGCGGTTATCTTTTTCGCCGTGCTGCATTCAGCCATTACACGG 561
                                **************************************************

sfsA_TOPO_pTrcHis.trimmed.seq   TTTTCACCCGCGCGCCACATCGATGAGAAATACGCGCAACTATTGTCAGA 850
EG10949                         TTTTCACCCGCGCGCCACATCGATGAGAAATACGCGCAACTATTGTCAGA 611
                                **************************************************

sfsA_TOPO_pTrcHis.trimmed.seq   AGCTCAACAGAGGGGGGTAGAAATTCTGGCTTACAAAGCGGAAATTTCTG 900
EG10949                         AGCTCAACAGAGGGGGGTAGAAATTCTGGCTTACAAAGCGGAAATTTCTG 661
                                **************************************************

sfsA_TOPO_pTrcHis.trimmed.seq   CTGAAGGCATGGCTCT-AAAAAATCACTGCCG------------ 931
EG10949                         CTGAAGGCATGGCTCTTAAAAAATCACTGCCGGTTACATTGTAG 705
                                *************** ***************


CLUSTAL W (1.82) multiple sequence alignment

REVERSED!!!!


yheO_TOPO_pTrcHis.trimmed.seq   GGGTCAGGTGGGACCACCGCGCTACTGCCGCCAGGCAAATTCTGTTTTAT 50
G7715                           --------------------------------------------------


yheO_TOPO_pTrcHis.trimmed.seq   CAGACCGCTTCTGCGTTCTGATTTAATCTGTATCAGGCTGAAAATCTTCT 100
G7715                           --------------------------------------------------


yheO_TOPO_pTrcHis.trimmed.seq   CTCATCCGCCAAAACAGCCAAGCTTCGAATTCGCCCTTGTGTTTTTTTTC 150
G7715                           ----------------------------------------GTGTTTTTTTC 12
                                                                        ***********

yheO_TOPO_pTrcHis.trimmed.seq   ATGTCCAGGTCGCTTTTAACCAACGAAACCAGTGAGTTGGATTTACTGGA 200
G7715                           ATGTCCAGGTCGCTTTTAACCAACGAAACCAGTGAGTTGGATTTACTGGA 62
                                **************************************************

yheO_TOPO_pTrcHis.trimmed.seq   TCAACGTCCTTTCGACCAGACCGATTTTGATATTCTGAAATCCTACGAAG 250
G7715                           TCAACGTCCTTTCGACCAGACCGATTTTGATATTCTGAAATCCTACGAAG 112
                                **************************************************

yheO_TOPO_pTrcHis.trimmed.seq   CGGTGGTGGACGGGTTAGCGATGCTTATTGGCTCCCACTGTGAAATCGTT 300
G7715                           CGGTGGTGGACGGGTTAGCGATGCTTATTGGCTCCCACTGTGAAATCGTT 162
                                **************************************************

yheO_TOPO_pTrcHis.trimmed.seq   TTGCACTCTTTGCAGGATCTAAAATGTTCAGCCATTCGCATTGCTAACGG 350
G7715                           TTGCACTCTTTGCAGGATCTAAAATGTTCAGCCATTCGCATTGCTAACGG 212
                                **************************************************

yheO_TOPO_pTrcHis.trimmed.seq   TGAACATACAGGCCGGAAGATTGGTTCGCCAATTACTGACCTGGCGCTAC 400
G7715                           TGAACATACAGGCCGGAAGATTGGTTCGCCAATTACTGACCTGGCGCTAC 262
                                **************************************************

yheO_TOPO_pTrcHis.trimmed.seq   GTATGCTGCACGATATGACGGGAGCGGATAGCAGCGTTTCTAAATGCTAC 450
G7715                           GTATGCTGCACGATATGACGGGAGCGGATAGCAGCGTTTCTAAATGCTAC 312
                                **************************************************

yheO_TOPO_pTrcHis.trimmed.seq   TTTACTCGCGCCAAAAGCGGCGTATTAATGAAGTCCCTGACTATCGCGAT 500
G7715                           TTTACTCGCGCCAAAAGCGGCGTATTAATGAAGTCCCTGACTATCGCGAT 362
                                **************************************************

yheO_TOPO_pTrcHis.trimmed.seq   TCGTAACCGCGAACAGCGTGCAATTGGTCTGCTGTGCATCAATATGAATC 550
G7715                           TCGTAACCGCGAACAGCGTGTAATTGGTCTGCTGTGCATCAATATGAATC 412
                                ****************** ******************************

yheO_TOPO_pTrcHis.trimmed.seq   TTGATGTTCCCTTCTCGCAGATTATGAGCACCTTTGTGCCGCCAGAAACC 600
G7715                           TTGATGTTCCCTTCTCGCAGATTATGAGCACCTTTGTGCCGCCAGAAACC 462
                                **************************************************

yheO_TOPO_pTrcHis.trimmed.seq   CCGGATGTCGGTTCAAGCGTCAACTTTGCCTCTTCTGTTGAAGATCTGGT 650
G7715                           CCGGATGTCGGTTCAAGCGTCAACTTTGCCTCTTCTGTTGAAGATCTGGT 512
                                **************************************************

yheO_TOPO_pTrcHis.trimmed.seq   TACCCAAACGCTGGAGTTCACCATCGAAGAAGTGAATGCCGATCGCAATG 700
G7715                           TACCCAAACGCTGGAGTTCACCATCGAAGAAGTGAATGCCGATCGCAATG 562
                                **************************************************

yheO_TOPO_pTrcHis.trimmed.seq   TTTCTAATAACGCCAAAAATCGTCAGATCGTGCTGAATCTCTACGAGAAA 750
G7715                           TTTCTAATAACGCCAAAAATCGTCAGATCGTGCTGAATCTCTACGAGAAA 612
                                **************************************************
```

```
yheO_TOPO_pTrcHis.trimmed.seq      GGGATCTTCGATATTAAAGATGCGATCAACCAGGTTGTTGACCGCCTGAA 800
G7715                              GGGATCTTCGATATTAAAGATGCGATCAACCAGGTTGCTGACCGCCTGAA 662
                                   ************************************ ************

yheO_TOPO_pTrcHis.trimmed.seq      CATCTCCAAACACACTGTCTATCTCTACATCCGCCAGTTCAAGAGCGGTG 850
G7715                              CATCTCCAAACACACTGTCTATCTCTACATCCGCCAGTTCAAGAGCGGTG 712
                                   **************************************************

yheO_TOPO_pTrcHis.trimmed.seq      ATTTCCAGGGGCAAGATAAGAAAAGGGTGGATCCTAT 887
G7715                              ATTTCCAGGGGCAAGATAAGTAA-------------- 735
                                   ******************* **


CLUSTAL W (1.82) multiple sequence alignment

REVERSED!!!!!


yhiF_TOPO_pTrcHis.trimmed.seq      TTCGCAACGTTCAAATCCGCTCCCGGGCGGATTTGTCCTACTCAGGAGAGC 50
EG11889                            --------------------------------------------------


yhiF_TOPO_pTrcHis.trimmed.seq      GTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTTCGACTGAG 100
EG11889                            --------------------------------------------------


yhiF_TOPO_pTrcHis.trimmed.seq      CCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCATGGGGAGA 150
EG11889                            --------------------------------------------------


yhiF_TOPO_pTrcHis.trimmed.seq      CCCCACACTACCATCGGCGCTACGGCGTTTCACTTCTGAGTTCGGCATGG 200
EG11889                            --------------------------------------------------


yhiF_TOPO_pTrcHis.trimmed.seq      GGTCAGGTGGGACCACCGCGCTACTGCCGCCAGGCAAATTCTGTTTTATC 250
EG11889                            --------------------------------------------------


yhiF_TOPO_pTrcHis.trimmed.seq      AGACCGCTTCTGCGTTCTGATTTAATCTGTATCAGGCTGAAAATCTTCTC 300
EG11889                            --------------------------------------------------


yhiF_TOPO_pTrcHis.trimmed.seq      TCATCCGCCAAAACAGCCAAGCTTCGAATTCGCCCTTATGTTTCTTATAA 350
EG11889                            ----------------------------------------ATGTTTCTTATAA 13
                                                                           ************

yhiF_TOPO_pTrcHis.trimmed.seq      TTACCAGGGATACGATGTTCTTCACCGCGATGAAAAACATTCTGAGTAAA 400
EG11889                            TTACCAGGGATACGATGTTCTTCACCGCGATGAAAAACATTCTGAGTAAA 63
                                   **************************************************

yhiF_TOPO_pTrcHis.trimmed.seq      GGTAATGTCGTTCATATACAGAACGAAGAAGAGATCGACGTAATGTTGCA 450
EG11889                            GGTAATGTCGTTCATATACAGAACGAAGAAGAGATCGACGTAATGTTGCA 113
                                   **************************************************

yhiF_TOPO_pTrcHis.trimmed.seq      TCAGAATGCCTTCGTCATTATTGATACATTAATGAATAATGTATTTCATT 500
EG11889                            TCAGAATGCCTTCGTCATTATTGATACATTAATGAATAATGTATTTCATT 163
                                   **************************************************

yhiF_TOPO_pTrcHis.trimmed.seq      CTAATTTTCTCACTCAAATTGAACGATTAAAACCTGTCCATGTCATTATT 550
EG11889                            CTAATTTTCTCACTCAAATTGAACGATTAAAACCTGTCCATGTCATTATT 213
                                   **************************************************

yhiF_TOPO_pTrcHis.trimmed.seq      TTCTCCCCCTTTAATATTAAACGCTGCCTGGGGAAAGTGCCGGTGACCTT 600
EG11889                            TTCTCCCCCTTTAATATTAAACGCTGCCTGGGGAAAGTGCCGGTGACCTT 263
                                   **************************************************

yhiF_TOPO_pTrcHis.trimmed.seq      TGTTCCGCGGACTATCACTATCATTGATTTTGTCGCACTCATCAATGGCA 650
EG11889                            TGTTCCGCGGACTATCACTATCATTGATTTTGTCGCACTCATCAATGGCA 313
                                   **************************************************

yhiF_TOPO_pTrcHis.trimmed.seq      GTTACTGCTCTGTGCCTGAAGCGGCTGTGTCACTTTCGCGCAAGCAACAT 700
EG11889                            GTTACTGCTCTGTGCCTGAAGCGGCTGTGTCACTTTCGCGCAAGCAACAT 363
                                   **************************************************

yhiF_TOPO_pTrcHis.trimmed.seq      CAGGTTCTGAGCTGCATTGCGAATCAAATGACAACGGAAGATATTCTGGA 750
EG11889                            CAGGTTCTGAGCTGCATTGCGAATCAAATGACAACGGAAGATATTCTGGA 413
                                   **************************************************

yhiF_TOPO_pTrcHis.trimmed.seq      GAAACTGAAAATATCGCTAAAAACGTTCTACTGCCATAAACACAATATCA 800
EG11889                            GAAACTGAAAATATCGCTAAAAACGTTCTACTGCCATAAACACAATATCA 463
                                   **************************************************

yhiF_TOPO_pTrcHis.trimmed.seq      TGATGATCCTCAATCTTAAGCGGATCAATGAGCTGGTACGCCATCAGCAT 850
EG11889                            TGATGATCCTCAATCTTAAGCGGATCAATGAGCTGGTACGCCATCAGCAT 513
                                   **************************************************

yhiF_TOPO_pTrcHis.trimmed.seq      ATTGATTATCTGGTGTGAAAGGGTTGATCCTAT 883
EG11889                            ATTGATTATCTGGTGTGA---------------- 531
                                   ******************
```

Brief Conclusions: The Qiagen prep definitely outperformed the Eppy kit. The yield was higher and the purity was better. Also, the Qiagen columns can use up to 5ml of culture, while the Eppy kit recommends 3ml max, so Qiagen's yield should be higher if I use more than 3ml of culture next time.

The sequencing results are horrible of the 19 sequencing reactions on 14 vectors, only 1 came back error-free. One error was my fault because I designed the primer wrong. Hopefully the next round will see improved results.

1.3.3 Recloning TOPO constructs

Out of the 15 or so sequences sent to Agencourt, all but one had errors. I'm going to reclone them using *Easy-A High-Fidelity PCR Master Mix*. Hopefully, the error rate will go down considerably. I'm going to prepare genomic DNA to PCR from this time, which should give a cleaner PCR.

Plated colonies from freezer stock: Mon Oct 31 10:45AM Need to make 40 LB:glucose:agar plates with amp according to the TOPO kit instructions.

Checking the imaging system on large gels

Mon Oct 27, 2005 In order to clone in higher-throughput I need to get the larger gel system working. With the CAB imaging system the resolution it was impossible to get high enough resolution for a good image. Plus, my loading dye was too concentrated and created a huge visual obstruction on the gel (see Fig 1.4).

Here I ran 2.5μl of each PCR product using 1/10 fisher dye in a 10μl total volume. Gel was increased to 300ml to make it thicker and more loadable.

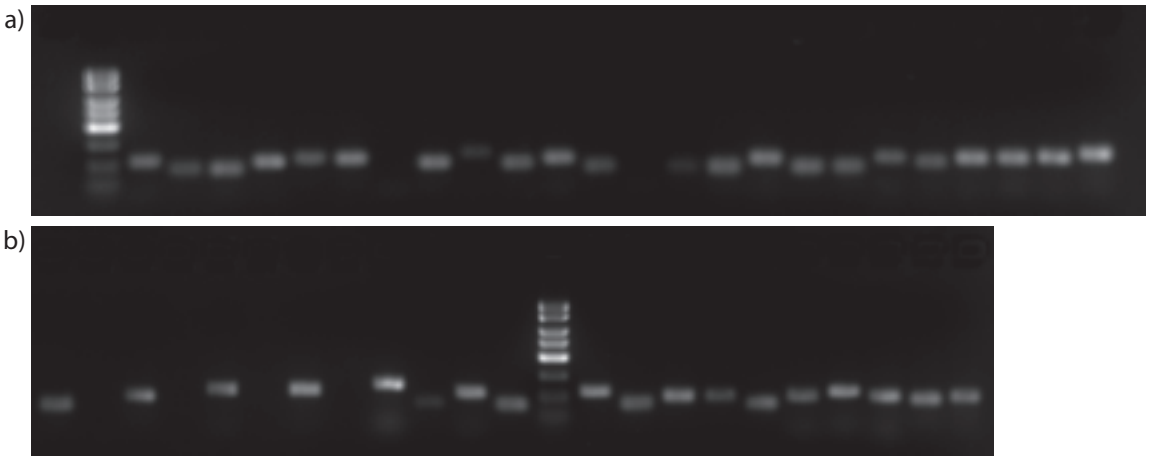


Figure 1.16: 300 ml, 1% agarose gel with 1.5 ul of 1% ethidium bromide run for 38 min at 120 volts. 2.5μl of each PCR sample was used.

Brief Conclusions: I'm still not totally satisfied with the look of the gel. For the TOPO system the framents will be longer which will make identification easier, but the resolution is still a bit crappy. I need to try using the zoom lense to see if this helps, as the bands themselves really are

quite small. Perhaps optics can help. Next time I should move too the 1/40 fisher dye as this 1/10 is still slightly in the way. Also, I made samples on parafilm. This is tricky with the multichannel.

Transfering oligos to 96-well plate

Wed Nov 2 11:08:18 EST 2005

Cloning will be done from plates. This sample size 36 is right on the border where I could do it by hand or by plate. Since every other step is on plates, I'm going to put in the extra effort to move the oligos to plates. This will make things easier to scale up later should I need to. The contents of the plate are shown in Table 1.3.3

TOPO TF cloning oligo plate												
-	1	2	3	4	5	6	7	8	9	10	11	12
A	bolA	cbl	fecI	flhC	gadX	hyaC	lexA*	nac	pdhR	rhaR	sfsA	yheO
B	yhiF	glcC	nusA	yhiW	ydaK	gfp*	-	bolA-m	cbl-m	fecI-m	flhC-m	gadX-m
C	hyaC-m	lexA-m*	lrp-m	nac-m	pdhR-m	rhaR-m	sfsA-m	yheO-m	yhiF-m	gfp-m	fliA	fliA-m
D	-	-	-	-	-	-	-	-	-	-	-	-
E	-	-	-	-	-	-	-	-	-	-	-	-
F	-	-	-	-	-	-	-	-	-	-	-	-
G	-	-	-	-	-	-	-	-	-	-	-	-
H	-	-	-	-	-	-	-	-	-	-	-	-

Table 1.2: Oligo concentration is 4μM. Wells contain both a forward and reverse primer. lexA contains the corrected lexA reverse primer. gfp requires a different template (since it is not in genome it must be amplified from plasmid). Oligo names with -m indicate the myc version of the primer that removes the stop codon (necessary for proper production of the myc tag). bolA-m didn't make it into the final plate because I forgot to order the primer.

Preparing Genomic DNA

Wed Nov 2 11:14:22 EST 2005

A plate of MG1655 was grown 9hrs and placed in the refridgerator for 1 day (because I had jury duty on Nov 1, 2005). Two colonies were inoculated at 10AM, Nov 2 into 4ml of LB. Genomic DNA will be prepared as described in C.5 using 4ml of sample. Yields are:

Gene	Amount of Vector (ng/μl)	260/280	260/230	μl for 75 ng
genomicA	678.2	2.02	2.03	0.1106
genomicB	444.7	2.13	2.17	0.1687

Brief Conclusions: It should be noted that no RNAse digestion was performed, some of the yield is RNA.

PCR

PCRs will be performed on all 33 genes in Table 1.3.3 using the following reaction:

PCR Reaction composition	
H_2O	20 μl
Easy-A Master Mix	25 μl
Forward and reverse primer	0.4 mM
template DNA	75 ng

Thermal cycler conditions		
Initial denaturation	5 min	95°C
3-Step cycling		
Denature:	30 sec	95°C
Anneal:	30 sec	58°C
Extend:	30 sec	72°C
Number of Cycles:	28	
Final Extention:	25 min	72°C

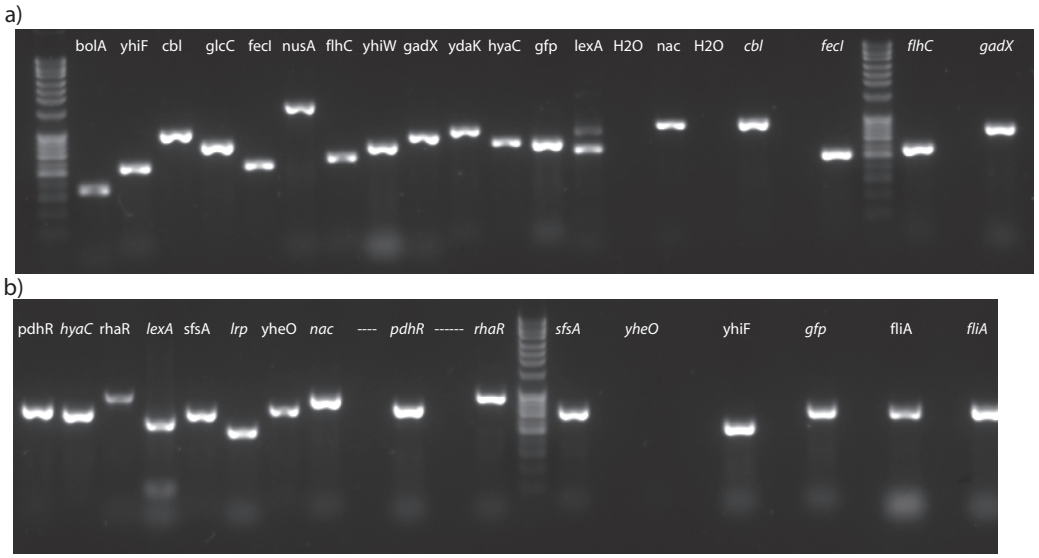


Figure 1.17: 300 ml, 1.5% agarose gel with 1.5 ul of 1% ethidium bromide run for 45 min at 120 volts. 5μl of each PCR sample was used. Sizes for the genes should be: bolA=351bp, cbl=951bp, fecI=521bp, fliA=719bp, flhC=579bp, gadX=825bp, hyaC=708bp, lexA=608bp, lrp=494bp, nac=917bp, pdhR=764bp, rhaR=939bp, sfsA=705bp, yheO=735bp, yhiF(dctR)=531bp, gfp=717bp, glcC=765bp, nusA=1488bp, yhiW=729bp, ydaK=909bp

Enough master mix for 39 reactions (3 extra) will be prepared, gfp will have additional template added by hand but will also contain genomic DNA. genomicA DNA was sticky (probably because of the high DNA conc?), so I'm going to use genomicB.

Brief Conclusions: There was just enough master mix in reservoir using 3 extra. These giant gels should really be run a little longer. All the PCRs looked great except the lexA which has a second band for some reason. Also, yheO-myc doesn't appear to have amplified.

Cloning

Thur Nov 3, 2005

0.75μl of each PCR product will be cloned into the appropriate pTrsHis TOPO vector according to the manual. Two concentrations of TOPO/insert mix are plated onto LB:agar:glucose:amp plates.

I didn't have enough pTrcHis2 vector for all of the genes I PCR'd, so yhiE and fliA were not cloned into the myc-tagged vector.

Plasmid prep

Fri Nov 4 17:34:17 EST 2005

3 colonies are chosen from each 96 total for miniprep (only two were chosen for yheO-myc. The matrix of picked colonies is:

Picked TOPO Colonies for Minipreping												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	pdhR $L\alpha\dagger$	pdhR $L\beta$	pdhR $H\alpha$	gadX $L\alpha$	gadX $L\beta\dagger$	gadX $H\alpha$	rhaR $L\alpha$	rhaR $H\alpha$	rhaR $H\beta$	fliA $L\alpha$	fliA $L\beta$	fliA $H\alpha$
B	sfsA $L\alpha$	sfsA $L\beta$	sfsA $H\alpha$	yheO $L\alpha$	yheO $L\beta$	yheO $H\alpha\dagger$	gfp $L\alpha\dagger$	gfp $L\beta$	gfp $H\alpha$	yhiF $L\alpha$	yhiF $L\beta\dagger$	yhiF $H\alpha$
C	hyaC $L\alpha$	hyaC $L\beta$	hyaC $H\alpha\dagger$	bolA $L\alpha$	bolA $L\beta$	bolA $H\alpha$	cbl $L\alpha$	cbl $L\beta\dagger$	cbl $H\alpha\dagger$	fecI $H\alpha$	fecI $H\beta$	fecI $H\gamma$
D	flhC $L\alpha$	flhC $L\beta$	flhC $H\alpha\dagger$	yhiW $L\alpha\dagger$	yhiW $L\beta$	yhiW $H\alpha$	nac $L\alpha$	nac $L\beta$	nac $H\alpha$	lexA $L\alpha$	lexA $H\alpha\dagger$	lexA $H\beta$
E	glcC $L\alpha$	glcC $L\beta$	glcC $H\alpha\dagger$	ydaK $L\alpha\dagger$	ydaK $L\beta$	ydaK $H\alpha$	nusA $L\alpha$	nusA $L\beta$	nusA $H\alpha$	cbl-m $L\alpha$	cbl-m $L\beta$	cbl-m $H\alpha\dagger$
F	hyaC-m $L\alpha$	hyaC-m $L\beta$	hyaC-m $H\alpha$	fecI-m $L\alpha$	fecI-m $L\beta$	fecI-m $H\alpha\dagger$	lexA-m $L\alpha$	lexA-m $L\beta\dagger$	lexA-m $H\alpha\dagger$	pdhR-m $L\alpha$	pdhR-m $L\beta$	pdhR-m $H\alpha$
G	sfsA-m $L\alpha$	sfsA-m $L\beta$	sfsA-m $H\alpha$	nac-m $L\alpha$	nac-m $L\beta$	nac-m $H\alpha$	rhaR-m $L\alpha$	rhaR-m $L\beta$	rhaR-m $H\alpha$	lrp-m $L\alpha$	lrp-m $L\beta$	lrp-m $H\alpha$
H	gfp-m $L\alpha\dagger$	gfp-m $L\beta$	gfp-m $H\alpha$	gadX-m $L\alpha$	gadX-m $L\beta$	gadX-m $H\alpha$	flhC-m $L\alpha$	flhC-m $L\beta$	flhC-m $H\alpha$	yheO-m $H\alpha\dagger$	yheO-m $H\beta\dagger$	-

Table 1.3: L = low conc. plate (plated 10μl of ligation reaction), H = high conc. plate (35μl of ligation), † = incorrect/no insert in digest (see Figure 1.18)

It took 3 hrs to pick all those colonies, and if I had had half a brain (as I will if I ever repeat this). I would have taken a μl of each 1200μl sample and added it to another plate to make a freezer stock from. Alas, I'm too stupid this time; next time I won't be. Random samples were taken to determine how much of each sample to use for digests to check insert size. H12 was a negative control containing only LB, and its yield is the lowest of the bunch in the table below. When I spun down the plate prior to lysing, it was the only well with no colonies. The samples yields are:

Sample	**DNA (ng/μl)**	**260/280**	**260/230**
C10	30.2	1.84	1.38
G1	32.3	1.77	1.40
E6	20.7	1.62	1.43
H10	20.5	1.79	1.38
B9	15.4	1.56	1.00
H5	14.9	1.67	1.47
H12	7.7	1.46	0.65

Brief Conclusions: I noticed later that, either my samples evaporated or spilled out randomly from the overnight or I wasn't consistent with my pipetting of the 1.2ml sample in each well. I'd really prefer something that would allow 2ml of sample. I should use the air permeable tape next time and the maximum 1.3ml sample if not a different plate that allows 2ml. Also, I didn't have the right attachment to the vacuum to allow me to use the Qiagen TurboFilter plate properly. I've ordered one.

Insert check

Mon Nov 7, 2005

8 μl of each vector was digested to check for an insert of the correct length using the following reactions:

pTrcHis

Restriction Digest	
vector/DNA	8 μl
EcoRI buffer	2 μl
BSA	2 μl
EcoRI enzyme	0.4 μl
NcoI enzyme	0.4 μl
H_2O	7 μl

pTrcHis2

Restriction Digest	
vector/DNA	8 μl
EcoRI buffer	2 μl
BSA	2 μl
EcoRI enzyme	0.4 μl
BamHI enzyme	0.4 μl
H_2O	7 μl

I got these parameters from an earlier digest (see section 1.3.2). Unfortunately they are backwards. The NcoI enzyme works on pTrcHis, it is just expensive and makes the insert 100bp longer. The BamHI doesn't have a cutter site in pTrcHis2, so none of those insert checks worked, and I need to redo them.

Tue Nov 8 20:06:40 EST 2005

I reran the 39 genes with a myc tag using the proper enzymes (NcoI and EcoRI); things are much better this time (see Figure 1.19).

Brief Conclusions: In Figure 1.18, it is clear that none of the myc tagged genes are possible to check, because I used the wrong enzymes. However, most of the pTrcHis genes have the correct insert (see Table 1.3.3 for which ones). Also, 8μl seems sufficient even though the miniprep yield was poor.

Sequencing

Tue Nov 8 12:55:47 EST 2005

Twenty plasmids are being sent for initial sequencing. Others will be done if the first twenty look good. A few genes are being sequenced 2x (two different vectors, not he same vector 2x).

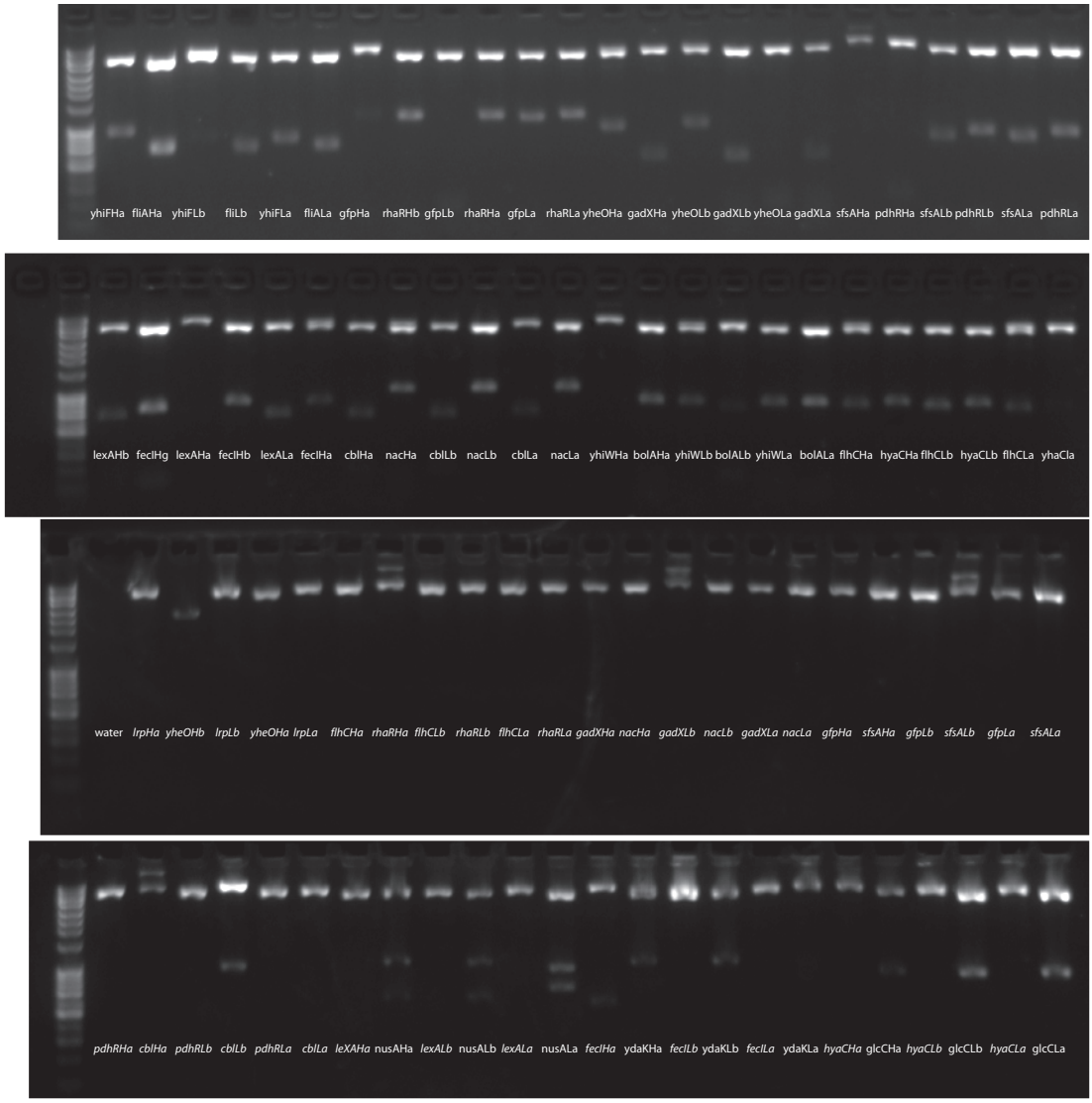


Figure 1.18: 300 ml, 1.8% agarose gel with 1.5 ul of 1% ethidium bromide run for 48 min at 120 volts. 20μl of each PCR sample was used.

Sample	DNA (ng/μl)	260/280	260/230	μl for 600ng
A2	28.0	1.68	1.00	21.42
A3	23.6	1.75	0.93	25.42
A4	50.9	1.58	0.88	11.79
A6	33.9	1.74	1.01	17.70
A7	18.0	1.84	1.54	33.33
A10	13.9	1.64	1.29	43.17
A11	70.2	1.58	0.79	8.54
B1	31.3	1.66	0.85	19.17
B4	18.8	1.30	1.29	31.91
B8	13.6	1.75	1.32	44.11
B9	15.4	1.56	1.00	38.96
B10	17.1	1.85	1.29	35.09
C1	29.1	1.92	1.69	20.62
C4	15.2	1.72	0.97	39.47
C7	22.5	1.76	1.33	26.67
C10	30.2	1.84	1.38	19.87
C11	32.1	1.72	1.14	18.69
D1	25.7	1.86	1.78	23.35
E1	32.1	1.89	1.77	18.69
E5	22.8	1.84	1.36	26.32

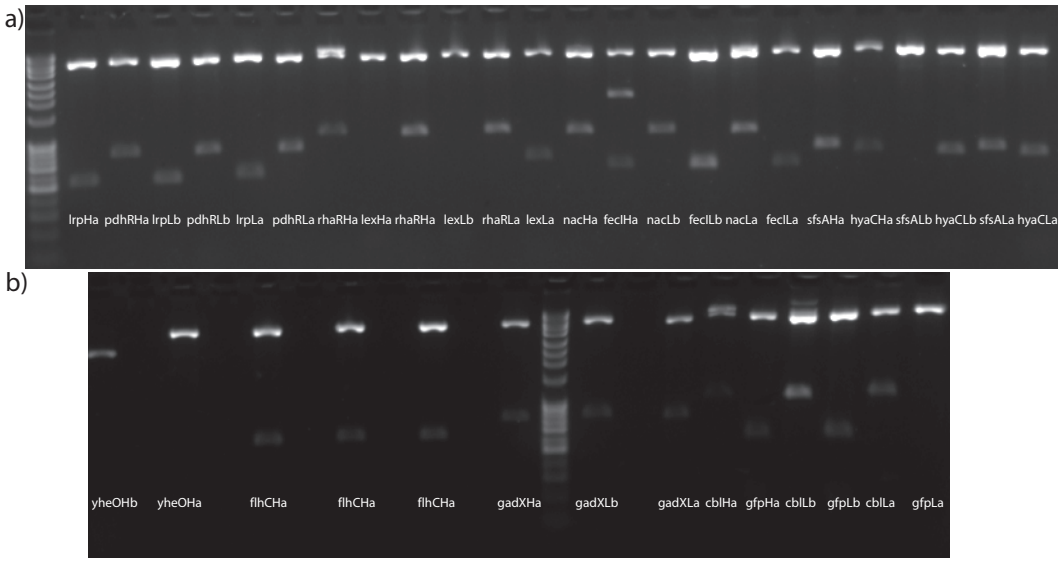


Figure 1.19: 300 ml, 1.8% agarose gel with 1.5 ul of 1% ethidium bromide run for 48 min at 120 volts. 20μl of each PCR sample was used.

The sequences with alignment to the known MG1655 sequence are:

```
CLUSTAL W (1.82) multiple sequence alignment


cbl_Trc_La.trimmed.seq      ATGACGATAGGATCCACCCTTGTGAATTTCCA-CAACTAAAGATAATCCG 49
G7071                       ----------------------GTGAATTTCCAACAACTAAAGATAATCCG 29
                                                   *********** *****************

cbl_Trc_La.trimmed.seq      CGAGGCTGCACGTCAGGATTACAACCTGACAGAGGTTGCGAATATGCTTT 99
G7071                       CGAGGCTGCACGTCAGGATTACAACCTGACAGAGGTTGCGAATATGCTTT 79
                            **************************************************

cbl_Trc_La.trimmed.seq      TTACCTCACAGTCAGGCGTCAGCCGTCATATTCGGGAACTGGAGGATGAA 149
G7071                       TTACCTCACAGTCAGGCGTCAGCCGTCATATTCGGGAACTGGAGGATGAA 129
                            **************************************************

cbl_Trc_La.trimmed.seq      CTTGGCATCGAAATATTTGTTCGACGAGGTAAGCGACTGCTGGGCATGAC 199
G7071                       CTTGGCATCGAAATATTTGTTCGACGAGGTAAGCGACTGCTGGGCATGAC 179
                            **************************************************

cbl_Trc_La.trimmed.seq      TGAACCGGGCAAAGCATTACTGGTCATTGCAGAACGTATTCTGAATGAAG 249
G7071                       TGAACCGGGCAAAGCATTACTGGTCATTGCAGAACGTATTCTGAATGAAG 229
                            **************************************************

cbl_Trc_La.trimmed.seq      CCAGTAATGTTCGTCGGCTTGCAGACCTGTTTACCAACGATACGTCTGGC 299
G7071                       CCAGTAATGTTCGTCGGCTTGCAGACCTGTTTACCAACGATACGTCTGGC 279
                            **************************************************

cbl_Trc_La.trimmed.seq      GTTCTCACTATTGCAACGACGCATACTCAGGCACGTTATAGCTTGCCAGA 349
G7071                       GTTCTCACTATTGCAACGACGCATACTCAGGCACGTTATAGCTTGCCAGA 329
                            **************************************************

cbl_Trc_La.trimmed.seq      GGTCATTAAAGCTTTTCGCGAACTTTTCCCGGAGGTTCGGCTCGAGCTAA 399
G7071                       GGTCATTAAAGCTTTTCGCGAACTTTTCCCGGAGGTTCGGCTCGAGCTAA 379
                            **************************************************

cbl_Trc_La.trimmed.seq      TCCAGGGGACGCCACAGGAAATTGCGACATTGTTGCAAAATGGCGAAGCT 449
G7071                       TCCAGGGGACGCCACAGGAAATTGCGACATTGTTGCAAAATGGCGAAGCT 429
                            **************************************************

cbl_Trc_La.trimmed.seq      GATATTGGTATCGCCAGCGAGCGTTTGAGTAATGACCCGCAGCTCGTCGC 499
G7071                       GATATTGGTATCGCCAGCGAGCGTTTGAGTAATGACCCGCAGCTCGTCGC 479
                            **************************************************

cbl_Trc_La.trimmed.seq      CTTCCCGTGGTTTCGTTGGCACCATAGTTTGCTTGTTCCACACGATCATC 549
G7071                       CTTCCCGTGGTTTCGTTGGCACCATAGTTTGCTTGTTCCACACGATCATC 529
                            **************************************************

cbl_Trc_La.trimmed.seq      CCTTGACGCAAATTTCACCATTGACGCTGGAATCAATAGCGAAGTGGCCG 599
G7071                       CCTTGACGCAAATTTCACCATTGACGCTGGAATCAATAGCGAAGTGGCCG 579
                            **************************************************
```

```
cbl_Trc_La.trimmed.seq      TTAATCACTTACCGACAGGGGATTACGGGGCGCTCACGTATTGATGACGC 649
G7071                       TTAATCACTTACCGACAGGGGATTACGGGGCGCTCACGTATTGATGACGC 629
                            **************************************************

cbl_Trc_La.trimmed.seq      ATTTGCCCGCAAAGGTTTGCTGGCAGATATTGTATTAAGTGCGCAGGATT 699
G7071                       ATTTGCCCGCAAAGGTTTGCTGGCAGATATTGTATTAAGTGCGCAGGATT 679
                            **************************************************

cbl_Trc_La.trimmed.seq      CTGATGTCATTAAAACCTATGTTGCTCTTGGGCTTGGGATCGGATTAGTT 749
G7071                       CTGATGTCATTAAAACCTATGTTGCTCTTGGGCTTGGGATCGGATTAGTT 729
                            **************************************************

cbl_Trc_La.trimmed.seq      GTCGAGCAATCCAGTGGCGAACAAGAGGAAGAGAAGAATTTAATCCGCCTGGA 799
G7071                       GCCGAGCAATCCAGTGGCGAACAAGAGGAAGAGAAGAATTTAATCCGCCTGGA 779
                            * **********************************************

cbl_Trc_La.trimmed.seq      TACGCGGCATCTTTTTGATGCTAATACTGTCTGGTTGGGACTGAAGCGAG 849
G7071                       TACGCGGCATCTTTTTGATGCTAATACTGTCTGGTTGGGACTGAAGCGAG 829
                            **************************************************

cbl_Trc_La.trimmed.seq      GACAACTTCAGCGTAACTATGTCTGGCGCTTTCTGGAACTTTGTAATGCA 899
G7071                       GACAACTTCAGCGTAACTATGTCTGGCGCTTTCTGGAACTTTGTAATGCA 879
                            **************************************************

cbl_Trc_La.trimmed.seq      AGACTGTCAGTTGAGGATATCAAGCGGC---------------------- 927
G7071                       GGACTGTCAGTTGAGGATATCAAGCGGCAGGTGATGGAAAGCAGTGAAGA 929
                             *************************

cbl_Trc_La.trimmed.seq      ----------------------
G7071                       GGAAATTGATTATCAGATATAG 951


CLUSTAL W (1.82) multiple sequence alignment

** BOTH SEQUENCES ARE BACKWARDS!!!!!

fecI                        --------------------------------------------------
fecI_Trc_Hb.trimmed.seq     AAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAG 50
fecI_Trc_Ha.trimmed.seq     --------------------------------------------------


fecI                        --------------------------------------------------
fecI_Trc_Hb.trimmed.seq     TTCCCTACTCTCGCATGGGGAGACCCCACACTACCATCGGCGCTACGGCG 100
fecI_Trc_Ha.trimmed.seq     --------------------------------------------------


fecI                        --------------------------------------------------
fecI_Trc_Hb.trimmed.seq     TTTCACTTCTGAGTTCGGCATGGGGTCAGGTGGGACCACCGCGCTACTGC 150
fecI_Trc_Ha.trimmed.seq     --------------------------------------------------


fecI                        --------------------------------------------------
fecI_Trc_Hb.trimmed.seq     CGCCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGATTTAATC 200
fecI_Trc_Ha.trimmed.seq     --------------------------------------------------


fecI                        --------------------------------------------------
fecI_Trc_Hb.trimmed.seq     TGTATCAGGCTGAAAATCTTCTCTCATCCGCCAAAACAGCCAAGCTTCGA 250
fecI_Trc_Ha.trimmed.seq     --------------------------------------------------


fecI                        ----------ATGTCTGACCGCGCCACTACCACAGCTTCCTTAACGTTCG 40
fecI_Trc_Hb.trimmed.seq     ATTCGCCCTTATGTCTGACCGCGCCACTACCACAGCTTCCTTAACGTTCG 300
fecI_Trc_Ha.trimmed.seq     ----------------GACCGCGCCACTACCACAGCTTCCTTAACGTTCG 34
                                            **********************************

fecI                        AGTCGCTTTATGGCACACATCACGGCTGGTTGAAAAGCTGGCTGACGCGC 90
fecI_Trc_Hb.trimmed.seq     AGTCGCTTTATGGCACACATCACGGCTGGTTGAAAAGCTGGCTGACGCGC 350
fecI_Trc_Ha.trimmed.seq     AGTCGCTTTATGGCACACATCACGGCTGGTTGAAAAGCTGGCTGACGCGC 84
                            **************************************************

fecI                        AAACTCCAGTCTGCTTTTGATGCAGATGACATTGCCCAGGACACTTTTTT 140
fecI_Trc_Hb.trimmed.seq     AAACTCCAGTCTGCTTTTGATGCAGATGACATTGCCCAGGACACTTTTTT 400
fecI_Trc_Ha.trimmed.seq     AAACTCCAGTCTGCTTTTGATGCAGATGACATTGCCCAGGACACTTTTTT 134
                            **************************************************

fecI                        GCGGGTAATGGTCAGCGAAACGCTCTCGACGATCCGCGATCCTCGCTCCT 190
fecI_Trc_Hb.trimmed.seq     GCGGGTAATGGTCAGCGAAACGCTCTCGACGATCCGCGATCCTCGCTCCT 450
fecI_Trc_Ha.trimmed.seq     GCGGGTAATGGTCAGCGAAACGCTCTCGACGATCCGCGATCCTCGCTCCT 184
                            **************************************************

fecI                        TCCTCTGCACTATCGCCAAACGCGTGATGGTGGACCTGTTTCGCCGAAAC 240
fecI_Trc_Hb.trimmed.seq     TCCTCTGCACTATCGCCAAACGCGTGATGGTGGACCTGTTTCGCCGAAAC 500
fecI_Trc_Ha.trimmed.seq     TCCTCTGCACTATCGCCAAACGCGTGATGGTGGACCTGTTTCGCCGAAAC 234
                            **************************************************

fecI                        GCGCTGGAAAAAGCGTATCTGGAGATGCTGGCGCTTATGCCGGAGGGGGG 290
fecI_Trc_Hb.trimmed.seq     GCGCTGGAAAAAGCGTATCTGGAGATGCTGGCGCTTATGCCGGAGGGGGG 550
fecI_Trc_Ha.trimmed.seq     GCGCTGGAAAAAGCGTATCTGGAGATGCTGGCGCTTATGCCGGAGGGGGG 284
                            **************************************************

fecI                        AGCGCCTTCACCTGAGGAACGCGAAAGCCAACTCGAGACCCTACAACTCC 340
```

```
fecI_Trc_Hb.trimmed.seq      AGCGCCTTCACCTGAGGAACGCGAAAGCCAACTCGAGACCCTACAACTCC 600
fecI_Trc_Ha.trimmed.seq      AGCGCCTTCACCTGAGGAACGCGAAAGCCAACTCGAGACCCTACAACTCC 334
                             **************************************************

fecI                         TCGACAGCATGCTGGACGGGCTAAACGGCAAAACACGTGAAGCGTTTCTG 390
fecI_Trc_Hb.trimmed.seq      TCGACAGCATGCTGGACGGGCTAAACGGCAAAACACGTGAAGCGTTTCTG 650
fecI_Trc_Ha.trimmed.seq      TCGACAGCATGCTGGACGGGCTAAACGGCAAAACACGTGAAGCGTTTCTG 384
                             **************************************************

fecI                         CTTTCGCAACTGGATGGTCTGACATACAGCGAGATTGCGCACAAACTCGG 440
fecI_Trc_Hb.trimmed.seq      CTTTCGCAACTGGATGGTCTGACATACAGCGAGATTGCGCACAAACTCGG 700
fecI_Trc_Ha.trimmed.seq      CTTTCGCAACTGGATGGTCTGACATACAGCGAGATTGCGCACAAACTCGG 434
                             **************************************************

fecI                         TGTTTCCATCAGCTCCGTGAAAAAATACGTGGCGAAAGCCGTCGAGCACT 490
fecI_Trc_Hb.trimmed.seq      TGTTTCCATCAGCTCCGTGAAAAAATACGTGGCGAAAGCCGTCGAGCACT 750
fecI_Trc_Ha.trimmed.seq      TGTTTCCATCAGCTCCGTGAAAAAATACGTGGCGAAAGCCGTCGAGCACT 484
                             **************************************************

fecI                         GCCTGCTGTTCCGTCTGGAGTATGGGTTATGA------------- 522
fecI_Trc_Hb.trimmed.seq      GCCTGCTGTTCCGTCTGGAGTATGGGTTATGAAAGGG-------- 787
fecI_Trc_Ha.trimmed.seq      GCCTGCTGTTCCGTCTGGAGTATGGGTTATGAAAGGGTTGGATCC 529
                             *********************************


CLUSTAL W (1.82) multiple sequence alignment

BACKWARDS


flhC_Trc_La.trimmed.seq      --------------------------------------------------
EG10319                      ATGAGTGAAAAAAGCATTGTTCAGGAAGCGCGGGATATTCAGCTGGCAAT 50


flhC_Trc_La.trimmed.seq      --------------------------------------------------
EG10319                      GGAATTGATCACCCTGGGCGCTCGTTTGCAGATGCTGGAAAGCGAAACAC 100


flhC_Trc_La.trimmed.seq      --------------------------------------------------
EG10319                      AGTTAAGTCGCGGACGCCTGATAAAACTTTATAAAGAACTGCGCGGAAGC 150


flhC_Trc_La.trimmed.seq      --------------------------------------------------
EG10319                      CCACCGCCGAAAGGCATGCTGCCATTCTCAACCGACTGGTTTATGACCTG 200


flhC_Trc_La.trimmed.seq      ----------TTCATGCTTCGATGTTCTGTAATGCATGGCAGTTTTTAC 39
EG10319                      GGAACAAAACGTTCATGCTTCGATGTTCTGTAATGCATGGCAGTTTTTAC 250
                                       ***************************************

flhC_Trc_La.trimmed.seq      TGAAAACCGGTTTGTGTAATGGCGTCGATGCGGTGATCAAAGCCTACCGT 89
EG10319                      TGAAAACCGGTTTGTGTAATGGCGTCGATGCGGTGATCAAAGCCTACCGT 300
                             **************************************************

flhC_Trc_La.trimmed.seq      TTATACCTTGAACAGTGCCCACAAGCAGAAGAAGGACCACTGCTGGCATT 139
EG10319                      TTATACCTTGAACAGTGCCCACAAGCAGAAGAAGGACCACTGCTGGCATT 350
                             **************************************************

flhC_Trc_La.trimmed.seq      AACCCGTGCCTGGACATTGGTGCGGTTTGTTGAAAGTGGATTACTGCAAC 189
EG10319                      AACCCGTGCCTGGACATTGGTGCGGTTTGTTGAAAGTGGATTACTGCAAC 400
                             **************************************************

flhC_Trc_La.trimmed.seq      TTTCCAGCTGCAACTGCTGCGGCGGCAATTTTATTACCCACGCTCACCAG 239
EG10319                      TTTCCAGCTGCAACTGCTGCGGCGGCAATTTTATTACCCACGCTCACCAG 450
                             **************************************************

flhC_Trc_La.trimmed.seq      CCTGTTGGCAGCTTTGCCTGCAGCTTATGTCAACCGCCATCCCGGGCAGT 289
EG10319                      CCTGTTGGCAGCTTTGCCTGCAGCTTATGTCAACCGCCATCCCGGGCAGT 500
                             **************************************************

flhC_Trc_La.trimmed.seq      AAAAAGACGTAAACTTTCCCAGAATCCTGCCGATATTATCCCACAACTGC 339
EG10319                      AAAAAGACGTAAACTTTCCCAGAATCCTGCCGATATTATCCCACAACTGC 550
                             **************************************************

flhC_Trc_La.trimmed.seq      TGGA-GAACAGAGAGTACAGGCTGTAAGGGTGGATCCTTATCGTCAT 385
EG10319                      TGGATGAACAGAGAGTACAGGCTGTTTAA------------------ 579
                             **** ********************


CLUSTAL W (1.82) multiple sequence alignment

BOTH BACKWARDS


EG11355                        --------------------------------------------------
fliA_Trc_E_Lb.trimmed.seq      GATTTAATCTGTATCAGGCTGAAAATCTTCTCTCATCCGCCAAAACAGCC 50
fliA_Trc_E_La.trimmed.seq      --------------------------------------------------


EG11355                        --------------------GTGAATTCACTCTATACCGCTGAAGGTGTAA 31
fliA_Trc_E_Lb.trimmed.seq      AAGCTTCGAATTCGCCCTTGTGAATTCACTCTATACCGCTGAAGGTGTAA 100
fliA_Trc_E_La.trimmed.seq      --------------------------------------------------
```

```
EG11355                         TGGATAAACACTCGCTGTGGCAGCGTTATGTCCCGCTGGTGCGTCACGAA 81
fliA_Trc_E_Lb.trimmed.seq       TGGATAAACACTCGCTGTGGCAGCGTTATGTCCCGCTGGTGCGTCACGAA 150
fliA_Trc_E_La.trimmed.seq       --------------------------------------------------


EG11355                         GCATTGCGCCTGCAGGTTCGACTGCCCGCGAGCGTGGAACTTGACGATCT 131
fliA_Trc_E_Lb.trimmed.seq       GCATTGCGCCTGCAGGTTCGACTGCCCGCGAGCGTGGAACTTGACGATCT 200
fliA_Trc_E_La.trimmed.seq       --------------------------------------------------


EG11355                         GCTACAGGCGGGCGGCATTGGGTTACTTAATGCCGTCGAACGCTATGACG 181
fliA_Trc_E_Lb.trimmed.seq       GCTACAGGCGGGCGGCATTGGGTTACTTAATGCCGTCGAACGCTATGACG 250
fliA_Trc_E_La.trimmed.seq       --------------------------------CCGTCGAACGATATGACG 18
                                                                 ********** *******

EG11355                         CCCTACAAGGAACGGCATTTACAACTTACGCAGTGCAGCGTATCCGTGGC 231
fliA_Trc_E_Lb.trimmed.seq       CCCTACAAGGAACGGCATTTACAACTTACGCAGTGCAGCGTATCCGTGGC 300
fliA_Trc_E_La.trimmed.seq       CCCTACAAGGAACGGCATTTACAACTTACGCAGTGCAGCGTATCCGTGGC 68
                                **************************************************

EG11355                         GCTATGCTGGATGAACTTCGCAGCCGTGACTGGGTGCCGCGCAGCGTGCG 281
fliA_Trc_E_Lb.trimmed.seq       GCTATGCTGGATGAACTTCGCAGCCGTGACTGGGTGCCGCGCAGCGTGCG 350
fliA_Trc_E_La.trimmed.seq       GCTATGCTGGATGAACTTCGCAGCCGTGACTGGGTGCCGCGCAGCGTGCG 118
                                **************************************************

EG11355                         ACGCAACGCGCGTGAAGTGGCACAGGCAATAGGGCAACTGGAGCAGGAAC 331
fliA_Trc_E_Lb.trimmed.seq       ACGCAACGCGCGTGAAGTGGCACAGGCAATAGGGCAACTGGAGCAGGAAC 400
fliA_Trc_E_La.trimmed.seq       ACGCAACGCGCGTGAAGTGGCACAGGCAATAGGGCAACTGGAGCAGGAAC 168
                                **************************************************

EG11355                         TTGGCCGCAACGCCACGGAAACTGAGGTAGCGGAACGTTTAGGGATCGAT 381
fliA_Trc_E_Lb.trimmed.seq       TTGGCCGCAACGCCACGGAAACTGAGGTAGCGGAACGTTTAGGGATCGAT 450
fliA_Trc_E_La.trimmed.seq       TTGGCCGCAACGCCACGGAAACTGAGGTAGCGGAACGTTTAGGGATCGAT 218
                                **************************************************

EG11355                         ATTGCCGATTATCGCCAAATGTTGCTCGACACCAATAACAGCCAGCTCTT 431
fliA_Trc_E_Lb.trimmed.seq       ATTGCCGATTATCGCCAAATGTTGCTCGACACCAATAACAGCCAGCTCTT 500
fliA_Trc_E_La.trimmed.seq       ATTGCCGATTATCGCCAAATGTTGCTCGACACCAATAACAGCCAGCTCTT 268
                                **************************************************

EG11355                         CTCCTACGATGAGTGGCGCGAAGAGCACGGCGATAGCATCGAACTGGTTA 481
fliA_Trc_E_Lb.trimmed.seq       CTCCTACGATGAGTGGCGCGAAGAGCACGGCGATAGCATCGAACTGGTTA 550
fliA_Trc_E_La.trimmed.seq       CTCCTACGATGAGTGGCGCGAAGAGCACGGCGATAGCATCGAACTGGTTA 318
                                **************************************************

EG11355                         CTGATGATCATCAGCGAGAAAACCCGCTACAACAACTACTGGACAGTAAT 531
fliA_Trc_E_Lb.trimmed.seq       CTGATGATCATCAGCGAGAAAACCCGCTACAACAACTACTGGACAGTAAT 600
fliA_Trc_E_La.trimmed.seq       CTGATGATCATCAGCGAGAAAACCCGCTACAACAACTACTGGACAGTAAT 368
                                **************************************************

EG11355                         CTGCGCCAGCGGGTGATGGAAGCCATCGAAACGTTGCCGGAGCGCGAAAA 581
fliA_Trc_E_Lb.trimmed.seq       CTGCGCCAGCGGGTGATGGAAGCCATCGAAACGTTGCCGGAGCGCGAAAA 650
fliA_Trc_E_La.trimmed.seq       CTGCGCCAGCGGGTGATGGAAGCCATCGAAACGTTGCCGGAGCGCGAAAA 418
                                **************************************************

EG11355                         ACTGGTATTAACCCTCTATTACCAGGAAGAGCTGAATCTCAAAGAGATTG 631
fliA_Trc_E_Lb.trimmed.seq       ACTGGTATTAACCCTCTATTACCAGGAAGAGCTGAATCTCAAAGAGATTG 700
fliA_Trc_E_La.trimmed.seq       ACTGGTATTAACCCTCTATTACCAGGAAGAGCTGAATCTCAAAGAGATTG 468
                                **************************************************

EG11355                         GCGCGGTGCTGGAGGTCGGGGAATCGCGGGTCAGTCAGTTACACAGCCAG 681
fliA_Trc_E_Lb.trimmed.seq       GCGCGGTGCTGGAGGTCGGGGAATCGCGGGTCAGTCAGTTACACAGCCAG 750
fliA_Trc_E_La.trimmed.seq       GCGCGGTGCTGGAGGTCGGGGAATCGCGGGTCAGTCAGTTACACAGCCAG 518
                                **************************************************

EG11355                         GCTATTAAACGGTTACGCACTAAACTGGGTAAGTTATAA---------- 720
fliA_Trc_E_Lb.trimmed.seq       GCTATTAAACGAATGCGCACTAAACTGGGTAAGTTATAAAAGGGT----- 795
fliA_Trc_E_La.trimmed.seq       GCTATTAAACGGTTACGCACTAAACTGGGTAAGTTATAAAAGGGTTGGAT 568
                                **********  * ************************

EG11355                         --
fliA_Trc_E_Lb.trimmed.seq       --
fliA_Trc_E_La.trimmed.seq       CC 570


CLUSTAL W (1.82) multiple sequence alignment

both BACKWARDS!!! :(


gadX_Trc_E_Ha.trimmed.seq       --------------------------TAATTGCGTATGCAAGACATAAAT 24
gadX_Trc_E_La.trimmed.seq       --------------------------------------------------
EG12243                         ATGCAATCACTACATGGGAATTGTCTAATTGCGTATGCAAGACATAAATA 50


gadX_Trc_E_Ha.trimmed.seq       ATATTCTCACCATGGTTAATGGTGAATATCGCTATTTTAATGGCGGTGAC 74
gadX_Trc_E_La.trimmed.seq       --------------------------------------------------
EG12243                         TATTCTCACCATGGTTAATGGTGAATATCGCTATTTTAATGGCGGTGACC 100
```

```
gadX_Trc_E_Ha.trimmed.seq      CTGGTTTTTGCGGATGCAAGCCAATTCGAGTAGATAAGTGTGTTGAAAAT 124
gadX_Trc_E_La.trimmed.seq      ---------------------------------------------------
EG12243                        TGGTTTTTGCGGATGCAAGCCAAATTCGAGTAGATAAGTGTGTTGAAAAT 150


gadX_Trc_E_Ha.trimmed.seq      TTTGTATTCGTGTCAAGGGACACGCTTTCATTATTTCTCCCGATGCTCAA 174
gadX_Trc_E_La.trimmed.seq      ---------------------------------------------------
EG12243                        TTTGTATTCGTGTCAAGGGACACGCTTTCATTATTTCTCCCGATGCTCAA 200


gadX_Trc_E_Ha.trimmed.seq      GGAGGAGGCATTAAATCTTCATGCACATAAAAAAGTTTCTTCATTACTCG 224
gadX_Trc_E_La.trimmed.seq      ---------------------------------------------------
EG12243                        GGAGGAGGCATTAAATCTTCATGCACATAAAAAAGTTTCTTCATTACTCG 250


gadX_Trc_E_Ha.trimmed.seq      TTCATCACTGTAGTAGAGATATTCCTGTTTTTCAGGAAGTTGCGCAACTA 274
gadX_Trc_E_La.trimmed.seq      ---------------------------------------------------
EG12243                        TTCATCACTGTAGTAGAGATATTCCTGTTTTTCAGGAAGTTGCGCAACTA 300


gadX_Trc_E_Ha.trimmed.seq      TCGCAGAATAAGAATCTTCGCTATGCGGAAATGCTACGTAAAAGAGCATT 324
gadX_Trc_E_La.trimmed.seq      ---------------------------------------------------
EG12243                        TCGCAGAATAAGAATCTTCGCTATGCAGAAATGCTACGTAAAAGAGCATT 350


gadX_Trc_E_Ha.trimmed.seq      AATCTTTGCGTTGTTATCTGTTTTTCTTGAGGATGAGCACTTTATACCGC 374
gadX_Trc_E_La.trimmed.seq      ------------GTTATCTGTTTTTCTGAGGGATGAGCACTTTATACCGC 38
EG12243                        AATCTTTGCGTTGTTATCTGTTTTTCTTGAGGATGAGCACTTTATACCGC 400
                                          ***************   ********************


gadX_Trc_E_Ha.trimmed.seq      TGCTTCTGAACGTTTTACAACCGAACATGCGAACACGAGTTTGTACGGTT 424
gadX_Trc_E_La.trimmed.seq      TGCTTCTGAACGTTTTACAACCGAACATGCGAACACGAGTTTGTACGGTT 88
EG12243                        TGCTTCTGAACGTTTTACAACCGAACATGCGAACACGAGTTTGTACGGTT 450
                               **************************************************


gadX_Trc_E_Ha.trimmed.seq      ATCAATAATAATATCGCCCATGAGTGGACACTAGCCCGAATCGCCAGCGA 474
gadX_Trc_E_La.trimmed.seq      ATCAATAATAATATCGCCCATGAGTGGACACTAGCCCGAATCGCCAGCGA 138
EG12243                        ATCAATAATAATATCGCCCATGAGTGGACACTAGCCCGAATCGCCAGCGA 500
                               **************************************************


gadX_Trc_E_Ha.trimmed.seq      GCTGTTGATGAGTCCAAGTCTGTTAAAGAAAAAATTGCGCGAAGAAGAGA 524
gadX_Trc_E_La.trimmed.seq      GCTGTTGATGAGTCCAAGTCTGTTAAAGAAAAAATTGCGCGAAGAAGAGA 188
EG12243                        GCTGTTGATGAGTCCAAGTCTGTTAAAGAAAAAATTGCGCGAAGAAGAGA 550
                               **************************************************


gadX_Trc_E_Ha.trimmed.seq      CATCATATTCACAGTTGCTTACTGAGTGTAGAATGCAACGTGCTTTGCAA 574
gadX_Trc_E_La.trimmed.seq      CATCATATTCACAGTTGCTTACTGAGTGTAGAATGCAACGTGCTTTGCAA 238
EG12243                        CATCATATTCACAGTTGCTTACTGAGTGTAGAATGCAACGTGCTTTGCAA 600
                               **************************************************


gadX_Trc_E_Ha.trimmed.seq      CTTATTGTTATACATGGTTTTTCAATTAAGCGAGTTGCAGTATCCTGTGG 624
gadX_Trc_E_La.trimmed.seq      CTTATTGTTATACATGGTTTTTCAATTAAGCGAGTTGCAGTATCCTGTGG 288
EG12243                        CTTATTGTTATACATGGTTTTTCAATTAAGCGAGTTGCAGTATCCTGTGG 650
                               **************************************************


gadX_Trc_E_Ha.trimmed.seq      ATATCACAGCGTGTCGTATTTCATTTACGTCTTTCGAAATTATTATGGGA 674
gadX_Trc_E_La.trimmed.seq      ATATCACAGCGTGTCGTATTTCATTTACGTCTTTCGAAATTATTATGGGA 338
EG12243                        ATATCACAGCGTGTCGTATTTCATTTACGTCTTTCGAAATTATTATGGGA 700
                               **************************************************


gadX_Trc_E_Ha.trimmed.seq      TGACGCCCACAGAGTATCAGGAGCGATCGGCGCAGAGATTGTCGAACCGT 724
gadX_Trc_E_La.trimmed.seq      TGACGCCCACAGAGTATCAGGAGCGATCGGCGCAGAGATTGTCGAACCGT 388
EG12243                        TGACGCCCACAGAGTATCAGGAGCGATCGGCGCAGAGATTGTCGAACCGT 750
                               **************************************************


gadX_Trc_E_Ha.trimmed.seq      GACTCGGCGGCAAGTATTGTTGCGCAAGGGAATTTTTACGGCACTGACCG 774
gadX_Trc_E_La.trimmed.seq      GACTCGGCGGCAAGTATTGTTGCGCAAGGGAATTTTTACGGCACTGACCG 438
EG12243                        GACTCGGCGGCAAGTATTGTTGCGCAAGGGAATTTTTACGGCACTGACCG 800
                               **************************************************


gadX_Trc_E_Ha.trimmed.seq      TTCTGCGGAAGGAATAAGATTATAGAAGGGTTGATCCTTAT----- 815
gadX_Trc_E_La.trimmed.seq      TTCTGCGGAAGGAATAAGATTATAGAAGGGTGGATCCTTATCGTCA 484
EG12243                        TTCTGCGGAAGGAATAAGATTATAG---------------------- 825
                               ************************


TAL W (1.82) multiple sequence alignment

BOTH BACKWARDS!!!!

gfp_Trc_Ha.trimmed.seq      CAGGTGGGACCACCGCGCTACTGCCGCCAGGCAAATTCTGTTTTATCAGA 50
Gfpmut3b                    --------------------------------------------------
gfp_Trc_Lb.trimmed.seq      ----------------------CCGCCAGGCAAATTCTGTTTTATCAGA 27


gfp_Trc_Ha.trimmed.seq      CCGCTTCTGCGTTCTGATNTAATCTGTATCAGGCTGAAAATCTTCTCTCA 100
Gfpmut3b                    --------------------------------------------------
gfp_Trc_Lb.trimmed.seq      CCGCTTCTGCGTTCTGATTTAATCTGTATCAGGCTGAAAATCTTCTCTCA 77


gfp_Trc_Ha.trimmed.seq      TCCGCCAAAACAGCCAAGCTTCGAATTCGCCCTTATGCGTAAAGGAGAAG 150
Gfpmut3b                    ----------------------------------ATGCGTAAAGGAGAAG 16
```

```
gfp_Trc_Lb.trimmed.seq      TCCGCCAAAACAGCCAAGCTTCGAATTCGCCCTTATGCGTAAAGGAGAAG 127
                                                        ***************

gfp_Trc_Ha.trimmed.seq      AACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTT 200
Gfpmut3b                    AACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTT 66
gfp_Trc_Lb.trimmed.seq      AACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTT 177
                            **************************************************

gfp_Trc_Ha.trimmed.seq      AATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATA 250
Gfpmut3b                    AATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATA 116
gfp_Trc_Lb.trimmed.seq      AATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATA 227
                            **************************************************

gfp_Trc_Ha.trimmed.seq      CGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTC 300
Gfpmut3b                    CGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTC 166
gfp_Trc_Lb.trimmed.seq      CGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTC 277
                            **************************************************

gfp_Trc_Ha.trimmed.seq      CATGGCCAACACTTGTCACTACTTTCGGTTATGGTGTTCAATGCTTTGCG 350
Gfpmut3b                    CATGGCCAACACTTGTCACTACTTTCGGTTATGGTGTTCAATGCTTTGCG 216
gfp_Trc_Lb.trimmed.seq      CATGGCCAACACTTGTCACTACTTTCGGTTATGGTGTTCAATGCTTTGCG 327
                            **************************************************

gfp_Trc_Ha.trimmed.seq      AGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGCCATGCC 400
Gfpmut3b                    AGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGCCATGCC 266
gfp_Trc_Lb.trimmed.seq      AGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGCCATGCC 377
                            **************************************************

gfp_Trc_Ha.trimmed.seq      CGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACT 450
Gfpmut3b                    CGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACT 316
gfp_Trc_Lb.trimmed.seq      CGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACT 427
                            **************************************************

gfp_Trc_Ha.trimmed.seq      ACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGA 500
Gfpmut3b                    ACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGA 366
gfp_Trc_Lb.trimmed.seq      ACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGA 477
                            **************************************************

gfp_Trc_Ha.trimmed.seq      ATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACA 550
Gfpmut3b                    ATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACA 416
gfp_Trc_Lb.trimmed.seq      ATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACA 527
                            **************************************************

gfp_Trc_Ha.trimmed.seq      CAAATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACA 600
Gfpmut3b                    CAAATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACA 466
gfp_Trc_Lb.trimmed.seq      CAAATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACA 577
                            **************************************************

gfp_Trc_Ha.trimmed.seq      AACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAA 650
Gfpmut3b                    AACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAA 516
gfp_Trc_Lb.trimmed.seq      AACAAAAGAATGGAATCAAGGTTAACTTCAAAATTAGACACAACATTGAA 627
                            ****************** ******************************

gfp_Trc_Ha.trimmed.seq      GATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGG 700
Gfpmut3b                    GATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGG 566
gfp_Trc_Lb.trimmed.seq      GATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGG 677
                            **************************************************

gfp_Trc_Ha.trimmed.seq      CGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTG 750
Gfpmut3b                    CGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTG 616
gfp_Trc_Lb.trimmed.seq      CGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTG 727
                            **************************************************

gfp_Trc_Ha.trimmed.seq      CCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAG 800
Gfpmut3b                    CCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAG 666
gfp_Trc_Lb.trimmed.seq      CCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAG 777
                            **************************************************

gfp_Trc_Ha.trimmed.seq      TTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATA 850
Gfpmut3b                    TTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATA 716
gfp_Trc_Lb.trimmed.seq      TTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATA 827
                            **************************************************

gfp_Trc_Ha.trimmed.seq      AAAGGGT------ 857
Gfpmut3b                    A------------ 717
gfp_Trc_Lb.trimmed.seq      AAAGGGTGGATCC 840
                            *


CLUSTAL W (1.82) multiple sequence alignment

Looks good, similar differences with TOPO like pdhR worries me

hyaC_Trc_La.trimmed.seq     ATGACGATAGGATCAACCCTTATGCAACAGAAAAGCGACAACGTTGTCAG 50
hyaC                        ----------------------ATGCAACAGAAAAGCGACAACGTTGTCAG 29
                                                   ****************************

hyaC_Trc_La.trimmed.seq     CCACTATGTCTTTGAAGCGCCAGTGCGCATCTGGCACTGGTTGACGGTGT 100
hyaC                        CCACTATGTCTTTGAAGCGCCAGTGCGCATCTGGCACTGGTTGACGGTGT 79
                            **************************************************

hyaC_Trc_La.trimmed.seq     TATGCATGGCGGTGTTGATGGTCACCGGATACTTTATCGGCAAGCCGCTA 150
```

```
hyaC                        TATGCATGGCGGTGTTGATGGTCACCGGATACTTTATCGGCAAGCCGCTA 129
                            **************************************************

hyaC_Trc_La.trimmed.seq     CCTTCCGTCAGCGGCGAGGCGACGTATCTGTTCTATATGGGCTACATCAG 200
hyaC                        CCTTCCGTCAGCGGCGAGGCGACGTATCTGTTCTATATGGGCTACATCAG 179
                            **************************************************

hyaC_Trc_La.trimmed.seq     GTTAATTCACTTCAGCGCCGGGATGGTTTTTACCGTGGTTTTGCTGATGC 250
hyaC                        GTTAATTCACTTCAGCGCCGGGATGGTTTTTACCGTGGTTTTGCTGATGC 229
                            **************************************************

hyaC_Trc_La.trimmed.seq     GGATCTACTGGGCTTTTGTTGGCAATCGATACTCCCGCGAGCTGTTTATC 300
hyaC                        GGATCTACTGGGCTTTTGTTGGCAATCGATACTCCCGCGAGCTGTTTATC 279
                            **************************************************

hyaC_Trc_La.trimmed.seq     GTGCCGGTATGGCGTAAAAGCTGGTGGCAGGGCGTGTGGTATGAAATCCG 350
hyaC                        GTGCCGGTATGGCGTAAAAGCTGGTGGCAGGGCGTGTGGTATGAAATCCG 329
                            **************************************************

hyaC_Trc_La.trimmed.seq     CTGGTATCTGTTTCTGGCAAAACGTCCGAGTGCCGATATAGGCCATAATC 400
hyaC                        CTGGTATCTGTTTCTGGCAAAACGTCCGAGTGCCGATATAGGCCATAATC 379
                            **************************************************

hyaC_Trc_La.trimmed.seq     CCATCGCCCAGGCGGCGATGTTCGGCTATTTCCTGATGTCGGTCTTTATG 450
hyaC                        CCATCGCCCAGGCGGCGATGTTCGGCTATTTCCTGATGTCGGTCTTTATG 429
                            **************************************************

hyaC_Trc_La.trimmed.seq     ATCATCACTGGTTTTGCGCTGTACAGCGAACACAGCCAGTACGCTATTTT 500
hyaC                        ATCATCACTGGTTTTGCGCTGTACAGCGAACACAGCCAGTACGCTATTTT 479
                            **************************************************

hyaC_Trc_La.trimmed.seq     TGCGCCGTTCCGTTATGTGGTGGAATTTTTCTACTGGACGGGTGGCAACT 550
hyaC                        TGCGCCGTTCCGTTATGTGGTGGAATTTTTCTACTGGACGGGTGGCAACT 529
                            **************************************************

hyaC_Trc_La.trimmed.seq     CAATGGACATTCACAGCTGGCATCGGCTGGGGATGTGGCTGATTGGCGCG 600
hyaC                        CAATGGACATTCACAGCTGGCATCGGCTGGGGATGTGGCTGATTGGCGCG 579
                            **************************************************

hyaC_Trc_La.trimmed.seq     TTTGTGATCGGTCATGTCTACATGGCGCTGCGTGAAGACATCATGTCCGA 650
hyaC                        TTTGTGATCGGTCATGTCTACATGGCGCTGCGTGAAGACATCATGTCCGA 629
                            **************************************************

hyaC_Trc_La.trimmed.seq     CGACACGGTGATCTCCACCATGGTCAACGGCTACCGTAGCCACAAATTTG 700
hyaC                        CGACACGGTGATCTCCACCATGGTCAACGGCTACCGTAGCCACAAATTTG 679
                            **************************************************

hyaC_Trc_La.trimmed.seq     GCAAAATAAGTAACAAGGAGCGTTCATGAAAGGGCGAATTCGAAGCTTGG 750
hyaC                        GCAAAATAAGTAACAAGGAGCGTTCATGA---------------------- 708
                            ****************************

hyaC_Trc_La.trimmed.seq     CTGTTTTGGCGGATGATAGAAGATTTTCAGCCTGATACAGATTAAATCAA 800
hyaC                        -------------------------------------------------


hyaC_Trc_La.trimmed.seq     AACGC 805
hyaC                        -----


CLUSTAL W (1.82) multiple sequence alignment

Lb is backwards
Ha looks good except some of the TOPO sequence look funky

pdhR                          --------------------------------------------------
pdhR_Trc_E_Lb.trimmed.seq     CTGATTTAATCTGTATCAGGCTGAAAATCTTCTCTCATCCGCCAAAACAG 50
pdhR_Trc_E_Ha.trimmed.seq     --------------------------------------------------


pdhR                          ----------------------ATGGCCTACAGCAAAATCCGCCAACCAAA 29
pdhR_Trc_E_Lb.trimmed.seq     CCAAGCTTCGAATTCGCCCTTATGGCCTACAGCAAAATCCGCCAACCAAA 100
pdhR_Trc_E_Ha.trimmed.seq     TGACGATAAGGATCCACCCTTATGGCCTACAGCAAAATCCGCCAACCAAA 50
                                                 ****************************

pdhR                          ACTCTCCGATGTGATTGAGCAGCAACTGGAGTTTTTGATCCTCGAAGGCA 79
pdhR_Trc_E_Lb.trimmed.seq     ACTCTCCGATGTGATTGAGCAGCAACTGGAGTTTTTGATCCTCGAAGGCA 150
pdhR_Trc_E_Ha.trimmed.seq     ACTCTCCGATGTGATTGAGCAGCAACTGGAGTTTTTGATCCTCGAAGGCA 100
                              **************************************************

pdhR                          CTCTCCGCCCCGGGCGAAAAACTCCCACCGGAACGCGAACTGGCAAAACAG 129
pdhR_Trc_E_Lb.trimmed.seq     CTCTCCGCCCGGGCGAAAAACTCCCACCGGAACGCGAACTGGCAAAACAG 200
pdhR_Trc_E_Ha.trimmed.seq     CTCTCCGCCCGGGCGAAAAACTCCCACCGGAACGCGAACTGGCAAAACAG 150
                              **************************************************

pdhR                          TTTGACGTCTCCCGTCCCTCCTTGCGTGAGGCGATTCAACGTCTCGAAGC 179
pdhR_Trc_E_Lb.trimmed.seq     TTTGACGTCTCCCGTCCCTCCTTGCGTGAGGCGATTCAACGTCTCGAAGC 250
pdhR_Trc_E_Ha.trimmed.seq     TTTGACGTCTCCCGTCCCTCCTTGCGTGAGGCGATTCAACGTCTCGAAGC 200
                              **************************************************

pdhR                          GAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTTTGTCCAGAGCA 229
pdhR_Trc_E_Lb.trimmed.seq     GAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTTTGTCCAGAGCA 300
pdhR_Trc_E_Ha.trimmed.seq     GAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTTTGTCCAGAGCA 250
```

```
                          **************************************************

pdhR                      GCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGCTCTCCGACCAT 279
pdhR_Trc_E_Lb.trimmed.seq GCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGCTCTCCGACCAT 350
pdhR_Trc_E_Ha.trimmed.seq GCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGCTCTCCGACCAT 300
                          **************************************************

pdhR                      CCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCCCTGGAAGGTAT 329
pdhR_Trc_E_Lb.trimmed.seq CCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCCCTGGAAGGTAT 400
pdhR_Trc_E_Ha.trimmed.seq CCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCCCTGGAAGGTAT 350
                          **************************************************

pdhR                      CGCCGCTTATTACGCCGCGCTGCGTAGTACCGATGAAGACAAGGAACGCA 379
pdhR_Trc_E_Lb.trimmed.seq CGCCGCTTATTACGCCGCGCTGCGTAGTACCGATGAAGACAAGGAACGCA 450
pdhR_Trc_E_Ha.trimmed.seq CGCCGCTTATTACGCCGCGCTGCGTAGTACCGATGAAGACAAGGAACGCA 400
                          **************************************************

pdhR                      TCCGTGAACTCCACCACGCCATAGAGCTGGCGCAGCAGTCTGGCGATCTG 429
pdhR_Trc_E_Lb.trimmed.seq TCCGTGAACTCCACCACGCCATAGAGCTGGCGCAGCAGTCTGGCGATCTG 500
pdhR_Trc_E_Ha.trimmed.seq TCCGTGAACTCCACCACGCCATAGAGCTGGCGCAGCAGTCTGGCGATCTG 450
                          **************************************************

pdhR                      GACGCGGAATCAAACGCCGTACTCCAGTATCAGATTGCCGTCACCGAAGC 479
pdhR_Trc_E_Lb.trimmed.seq GACGCGGAATCAAACGCCGTACTCCAGTATCAGATTGCCGTCACCGAAGC 550
pdhR_Trc_E_Ha.trimmed.seq GACGCGGAATCAAACGCCGTACTCCAGTATCAGATTGCCGTCACCGAAGC 500
                          **************************************************

pdhR                      GGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTATGGAGCCGATGT 529
pdhR_Trc_E_Lb.trimmed.seq GGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTATGGAGCCGATGT 600
pdhR_Trc_E_Ha.trimmed.seq GGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTATGGAGCCGATGT 550
                          **************************************************

pdhR                      TGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATTCGCGTCGCGAG 579
pdhR_Trc_E_Lb.trimmed.seq TGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATTCGCGTCGCGAG 650
pdhR_Trc_E_Ha.trimmed.seq TGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATTCGCGTCGCGAG 600
                          **************************************************

pdhR                      ATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTTGAAGCGATTAT 629
pdhR_Trc_E_Lb.trimmed.seq ATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTTGAAGCGATTAT 700
pdhR_Trc_E_Ha.trimmed.seq ATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTTGAAGCGATTAT 650
                          **************************************************

pdhR                      GGCCGGTAAGCCGGAAGAAGCGCGCGAAGCATCGCATCGCCATCTGGCCT 679
pdhR_Trc_E_Lb.trimmed.seq GGCCGGTAAGCCGGAAGAAGCGCGCGAAGCATCGCATCGCCATCTGGCCT 750
pdhR_Trc_E_Ha.trimmed.seq GGCCGGTAACCCGGAAGAAGCGCGCGAAGCATCGCATCGCCATCTGGCCT 700
                          ********* ******************************************

pdhR                      TTATCGAAGAAATTTTGCTCGACAGAAGTCGTGAAGAGAGCCGCCGTGAG 729
pdhR_Trc_E_Lb.trimmed.seq TTATCGAAGAAATTTTGCTCGACAGAAGTCGTGAAGAGAGCCGCCGTGAG 800
pdhR_Trc_E_Ha.trimmed.seq TTATCGAAGAAATTTTGCTCGACAGAAGTCGTGAAGAGAGCCGCCGTGAG 750
                          **************************************************

pdhR                      CGTTCTCTGCGTCGTCTGGAGCAACGAAAGAATTAG--------------- 765
pdhR_Trc_E_Lb.trimmed.seq CGTTCTCTGCGTCGTCTGGAGCAACGAAAGAATTAGAAGGG--------- 841
pdhR_Trc_E_Ha.trimmed.seq CGTTCTCTGCGTCGTCTGGAGCAACGAAAGAATTAGAAGGGCGAATTCGA 800
                          ***********************************

pdhR                      --------------------------------------------------
pdhR_Trc_E_Lb.trimmed.seq --------------------------------------------------
pdhR_Trc_E_Ha.trimmed.seq AGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATT 850


pdhR                      ---------------------------------
pdhR_Trc_E_Lb.trimmed.seq ---------------------------------
pdhR_Trc_E_Ha.trimmed.seq AAATCAGAACGCAGAAGCGGTCTGATAAAACA 882


CLUSTAL W (1.82) multiple sequence alignment

The deletion actually looks good on the chromatagraph, need to sequence the other way.


rhaR_Trc_E_La.trimmed.seq CCCTTATGGCTTTCTGCA-TAACGCGAATCTTCTCAACGTATTTGTACGC 49
EG10842                   -----ATGGCTTTCTGCAATAACGCGAATCTTCTCAACGTATTTGTACGC 45
                               ************ ******************************

rhaR_Trc_E_La.trimmed.seq CATATTGCGAATAATCAACTTCGTTCTCTGGCCGAGGTAGCCACGGTGGC 99
EG10842                   CATATTGCGAATAATCAACTTCGTTCTCTGGCCGAGGTAGCCACGGTGGC 95
                          **************************************************

rhaR_Trc_E_La.trimmed.seq GCATCAGTTAAAACTTCTCAAAGATGATTTTTTTGCCAGCGACCAGCAGG 149
EG10842                   GCATCAGTTAAAACTTCTCAAAGATGATTTTTTTGCCAGCGACCAGCAGG 145
                          **************************************************

rhaR_Trc_E_La.trimmed.seq CAGTCGCTGTGGCTGACCGTTATCCGCAAGATGTCTTTGCTGAACATACA 199
EG10842                   CAGTCGCTGTGGCTGACCGTTATCCGCAAGATGTCTTTGCTGAACATACA 195
                          **************************************************

rhaR_Trc_E_La.trimmed.seq CATGATTTTTGTGAGCTGGTGATTGTCTGGCGCGGTAATGGCCTGCATGT 249
EG10842                   CATGATTTTTGTGAGCTGGTGATTGTCTGGCGCGGTAATGGCCTGCATGT 245
                          **************************************************
```

```
rhaR_Trc_E_La.trimmed.seq      ACTCAACGATCGCCCTTATCGCATTACCCGTGGCGATCTCTTTTACATTC 299
EG10842                        ACTCAACGATCGCCCTTATCGCATTACCCGTGGCGATCTCTTTTACATTC 295
                               **************************************************

rhaR_Trc_E_La.trimmed.seq      ATGCTGACGATAAACACTCCTACGCTTCCGTTAACGATCTGGTTTTGCAG 349
EG10842                        ATGCTGACGATAAACACTCCTACGCTTCCGTTAACGATCTGGTTTTGCAG 345
                               **************************************************

rhaR_Trc_E_La.trimmed.seq      AATATTATTTATTGCCCGGAGCGTCTGAAGCTGAATCTTGACTGGCAGGG 399
EG10842                        AATATTATTTATTGCCCGGAGCGTCTGAAGCTGAATCTTGACTGGCAGGG 395
                               **************************************************

rhaR_Trc_E_La.trimmed.seq      GGCGATTCCGGGATTTAACGCCAGCGCAGGGCAACCACACTGGCGCTTAG 449
EG10842                        GGCGATTCCGGGATTTAACGCCAGCGCAGGGCAACCACACTGGCGCTTAG 445
                               **************************************************

rhaR_Trc_E_La.trimmed.seq      GTAGCATGGGGATGGCGCAGGCGCGGCAGGTTATCGGTCAGCTTGAGCAT 499
EG10842                        GTAGCATGGGGATGGCGCAGGCGCGGCAGGTTATCGGTCAGCTTGAGCAT 495
                               **************************************************

rhaR_Trc_E_La.trimmed.seq      GAAAGTAGTCAGCATGTGCCGTTTGCTAACGAAATGGCTGAGTTGCTGTT 549
EG10842                        GAAAGTAGTCAGCATGTGCCGTTTGCTAACGAAATGGCTGAGTTGCTGTT 545
                               **************************************************

rhaR_Trc_E_La.trimmed.seq      CGGGCAGTTGGTGATGTCGCTGA---------------------------- 572
EG10842                        CGGGCAGTTGGTGATGTTGCTGAATCGCCATCGTTACACCAGTGATTCGT 595
                               ***************** *****

rhaR_Trc_E_La.trimmed.seq      --------------------------------------------------
EG10842                        TGCCGCCAACATCCAGCGAAACGTTGCTGGATAAGCTGATTACCCGGCTG 645


rhaR_Trc_E_La.trimmed.seq      --------------------------------------------------
EG10842                        GCGGCTAGCCTGAAAAGTCCCTTTGCGCTGGATAAATTTTGTGATGAGGC 695


rhaR_Trc_E_La.trimmed.seq      --------------------------------------------------
EG10842                        ATCGTGCAGTGAGCGCGTTTTGCGTCAGCAATTTCGCCAGCAGACTGGAA 745


rhaR_Trc_E_La.trimmed.seq      --------------------------------------------------
EG10842                        TGACCATCAATCAATATCTGCGACAGGTCAGAGTGTGTCATGCGCAATAT 795


rhaR_Trc_E_La.trimmed.seq      --------------------------------------------------
EG10842                        CTTCTCCAGCATAGCCGCCTGTTAATCAGTGATATTTCGACCGAATGTGG 845


rhaR_Trc_E_La.trimmed.seq      --------------------------------------------------
EG10842                        CTTTGAAGATAGTAACTATTTTTCGGTGGTGTTTACCCGGGAAACCGGGA 895


rhaR_Trc_E_La.trimmed.seq      ---------------------------------------------
EG10842                        TGACGCCCAGCCAGTGGCGTCATCTCAATTCGCAGAAAGATTAA 939


CLUSTAL W (1.82) multiple sequence alignment

BACKWARDS

sfsA_Trc_E_La.trimmed.seq      TATCAGGCTGAAAAATCTTCTCTCATCCGCCAAAACAGCCAAGCTTCGAA 50
sfsA                           --------------------------------------------------


sfsA_Trc_E_La.trimmed.seq      TTCGCCCTTATGGAATTTTCTCCCCCTCTACAGCGCGCGACGCTAATTCA 100
sfsA                           ---------ATGGAATTTTCTCCCCCTCTACAGCGCGCGACGCTAATTCA 41
                                        ******************************************

sfsA_Trc_E_La.trimmed.seq      GCGTTACAAACGTTTTTTAGCCGATGTGATCACACCCGATGGTCGCGAAT 150
sfsA                           GCGTTACAAACGTTTTTTAGCCGATGTGATCACACCCGATGGTCGCGAAT 91
                               **************************************************

sfsA_Trc_E_La.trimmed.seq      TAACGCTACACTGCCCGAATACGGGTGCGATGACCGGTTGTGCAACGCCT 200
sfsA                           TAACGCTACACTGCCCGAATACGGGTGCGATGACCGGTTGTGCAACGCCT 141
                               **************************************************

sfsA_Trc_E_La.trimmed.seq      GGCGATACCGTCTGGTATTCGACTTCAGACAACACCAAACGGAAATACCC 250
sfsA                           GGCGATACCGTCTGGTATTCGACTTCAGACAACACCAAACGGAAATACCC 191
                               **************************************************

sfsA_Trc_E_La.trimmed.seq      ACACACCTGGGAATTAACTCAAAGCCAGAGCGGCGCATTTATTTGCGTCA 300
sfsA                           ACACACCTGGGAATTAACTCAAAGCCAGAGCGGCGCATTTATTTGCGTCA 241
                               **************************************************

sfsA_Trc_E_La.trimmed.seq      ACACGCTTTGGGCTGACAGGTTGACGAAAGAGGCTATCCTTAATGAATCA 350
sfsA                           ACACGCTTTGGGCTAACAGGTTGACGAAAGAGGCTATCCTTAATGAATCA 291
                               ************** *************************************

sfsA_Trc_E_La.trimmed.seq      ATTTCAGAACTGTCAGGCTATAGCTCGCTGAAAAGCGAAGTAAAATACGG 400
sfsA                           ATTTCAGAACTGTCAGGCTATAGCTCGCTGAAAAGCGAAGTAAAATACGG 341
                               **************************************************
```

```
sfsA_Trc_E_La.trimmed.seq      CGCAGAACGCAGCCGTATTGACTTTATGTTGCAGGCGGATTCGCGTCCAG 450
sfsA                           CGCAGAACGCAGCCGTATTGACTTTATGTTGCAGGCGGATTCGCGTCCAG 391
                               **************************************************

sfsA_Trc_E_La.trimmed.seq      ACTGCTATATTGAAGTGAAATCGGTTACGTTAGCGGAGAACGAACAGGGA 500
sfsA                           ACTGCTATATTGAAGTGAAATCGGTTACGTTAGCGGAGAACGAACAGGGA 441
                               **************************************************

sfsA_Trc_E_La.trimmed.seq      TATTTTCCCGATGCGGTCACTGAACGAGGTCAGAAACACCTTCGGGAGTT 550
sfsA                           TATTTTCCCGATGCGGTCACTGAACGAGGTCAGAAACACCTTCGGGAGTT 491
                               **************************************************

sfsA_Trc_E_La.trimmed.seq      GATGAGCGTAGCGGCTGAAGGCCAGCGTGCGGTTATCTTTTTCGCCGTGC 600
sfsA                           GATGAGCGTAGCGGCTGAAGGCCAGCGTGCGGTTATCTTTTTCGCCGTGC 541
                               **************************************************

sfsA_Trc_E_La.trimmed.seq      TGCATTCAGCCATTACACGGTTTTCACCCGCGCGCCCACATCGATGAGAAA 650
sfsA                           TGCATTCAGCCATTACACGGTTTTCACCCGCGCGCCCACATCGATGAGAAA 591
                               **************************************************

sfsA_Trc_E_La.trimmed.seq      TACGCGCAACTATTGTCAGAAGCTCAACAGAGGGGGGTAGAAATTCTGGC 700
sfsA                           TACGCGCAACTATTGTCAGAAGCTCAACAGAGGGGGGTAGAAATTCTGGC 641
                               **************************************************

sfsA_Trc_E_La.trimmed.seq      TTACAAAGCGGAAATTTCTGCTGAAGGCATGGCTCTTAAAAAATCACTGC 750
sfsA                           TTACAAAGCGGAAATTTCTGCTGAAGGCATGGCTCTTAAAAAATCACTGC 691
                               **************************************************

sfsA_Trc_E_La.trimmed.seq      CGGTTACATTGTAGAAGGGTGGATCC 776
sfsA                           CGGTTACATTGTAG------------ 705
                               **************
```

```
CLUSTAL W (1.82) multiple sequence alignment

Lb is BACKWARDS!
```

```
ydaK_Trc_Lb.trimmed.seq      ---------------------------------------------------
abgR                         ATGGCTTTTCAGGTAAAAATTCATCAAATTCGGGCTTTTGTTGAAGTGGC 50


ydaK_Trc_Lb.trimmed.seq      ---------------------------------------------------
abgR                         TCGTCAGGGCAGCATTCGCGGAGCGAGCCGAATGTTGAATATGTCGCAAC 100


ydaK_Trc_Lb.trimmed.seq      ------------------------GAGCTAGAAGAAGGGTTAGCGGCGCAA 27
abgR                         CGGCACTGAGTAAATCTATTCAGGAGCTAGAAGAAGGGTTAGCGGCGCAA 150
                                                     **************************

ydaK_Trc_Lb.trimmed.seq      CTCTTTTTTCGCCGTAGTAAAGGCGTGACGTTAACTGATGCCGGTGAAAG 77
abgR                         CTCTTTTTTCGCCGTAGTAAAGGCGTGACGTTAACTGATGCCGGTGAAAG 200
                             **************************************************

ydaK_Trc_Lb.trimmed.seq      TTTTTATCAGCACGCCAGTCTAATTCTTGAAGAGCTGCGCGCAGCCCAAG 127
abgR                         TTTTTATCAGCACGCCAGTCTAATTCTTGAAGAGCTGCGCGCAGCCCAAG 250
                             **************************************************

ydaK_Trc_Lb.trimmed.seq      AGGATATTCGCCAACGACAAGGGCAACTGGCAGGGCAGATTAATATCGGC 177
abgR                         AGGATATTCGCCAACGACAAGGGCAACTGGCAGGGCAGATTAATATCGGC 300
                             **************************************************

ydaK_Trc_Lb.trimmed.seq      ATGGGGGCCAGTATTTCCCGCAGTCTGATGCCAGCTGTCATATCTCGTTT 227
abgR                         ATGGGGGCCAGTATTTCCCGCAGTCTGATGCCAGCTGTCATATCTCGTTT 350
                             **************************************************

ydaK_Trc_Lb.trimmed.seq      TCATCAGCAGCATCCGCAGGTAAAAGTACGCATTATGGAAGGGCAACTGG 277
abgR                         TCATCAGCAGCATCCGCAGGTAAAAGTACGCATTATGGAAGGGCAACTGG 400
                             **************************************************

ydaK_Trc_Lb.trimmed.seq      TGTCGATGATTAATGAATTGCGTCAGGGAGAATTGGATTTCACCATCAAT 327
abgR                         TGTCGATGATTAATGAATTGCGTCAGGGAGAATTGGATTTCACCATCAAT 450
                             **************************************************

ydaK_Trc_Lb.trimmed.seq      ACCTATTATCAGGGACCGTACGACCACGAATTTACTTTTGAGAAATTACT 377
abgR                         ACCTATTATCAGGGACCGTACGACCACGAATTTACTTTTGAGAAATTACT 500
                             **************************************************

ydaK_Trc_Lb.trimmed.seq      GGAAAAGCAATTCGCGATCTTTTGCCGCCCGGGACACCCCGCCATTGGTG 427
abgR                         GGAAAAGCAATTCGCGATCTTTTGCCGCCCGGGACACCCCGCCATTGGTG 550
                             **************************************************

ydaK_Trc_Lb.trimmed.seq      CCCGTTCGATCAAACAGTTACTGGATTACAGCTGGACAATGCCGACGCCA 477
abgR                         CCCGTTCGATCAAACAGTTACTGGATTACAGCTGGACAATGCCGACGCCA 600
                             **************************************************

ydaK_Trc_Lb.trimmed.seq      CACGGCAGCTACTACAAACAGTTGAGTGAATTGCTTGACGATCAGGCGCA 527
abgR                         CACGGCAGCTACTACAAACAGTTGAGTGAATTGCTTGACGATCAGGCGCA 650
                             **************************************************

ydaK_Trc_Lb.trimmed.seq      AACGCCACAGGTCGGTGTAGTCTGCGAGACGTTCTCAGCCTGTATCAGTC 577
abgR                         AACGCCACAGGTCGGTGTAGTCTGCGAGACGTTCTCAGCCTGTATCAGTC 700
```

```
                          **************************************************

ydaK_Trc_Lb.trimmed.seq   TGGTGGCAAAAAGCGATTTTCTCAGCAAACTGCCTGAAGAAATGGGCTGC 627
abgR                      TGGTGGCAAAAAGCGATTTTCTCAGCAAACTGCCTGAAGAAATGGGCTGC 750
                          **************************************************

ydaK_Trc_Lb.trimmed.seq   GATCCCTTGCACGGACAGGGGCTGGTGATGTTGCCGGTTAGCGAAATTTT 677
abgR                      GATCCCTTGCACGGACAGGGGCTGGTGATGTTGCCGGTTAGCGAAATTTT 800
                          **************************************************

ydaK_Trc_Lb.trimmed.seq   ACCGAAAGCGGCCTATTATTTGATTCAGCGGCGTGATAGTCGCCAGACAC 727
abgR                      ACCGAAAGCGGCCTATTATTTGATTCAGCGGCGTGATAGTCGCCAGACAC 850
                          **************************************************

ydaK_Trc_Lb.trimmed.seq   CACTGACCGCGTCATTAATCACGCAATTCCGGCGAGAATGCGGCTATCTG 777
abgR                      CACTGACCGCGTCATTAATCACGCAATTCCGGCGAGAATGCGGCTATCTG 900
                          **************************************************

ydaK_Trc_Lb.trimmed.seq   CAAAGTTAAAGGGTGGATCCTTAT 801
abgR                      CAAAGTTAA--------------- 909
                          *********

CLUSTAL W (1.82) multiple sequence alignment

BACKWARDS


G7715                       GTGTTTTTTTTCATGTCCAGGTCGCTTTTAACCAACGAAACCAGTGAGTT 50
yheO_Trc_E_La.trimmed.seq   --------------------------------------------------


G7715                       GGATTTACTGGATCAACGTCCTTTCGACCAGACCGATTTTGATATTCTGA 100
yheO_Trc_E_La.trimmed.seq   --------------------------------------------------


G7715                       AATCCTACGAAGCGGTGGTGGACGGGTTAGCGATGCTTATTGGCTCCCAC 150
yheO_Trc_E_La.trimmed.seq   --------------------------------------------------


G7715                       TGTGAAATCGTTTTGCACTCTTTGCAGGATCTAAAATGTTCAGCCATTCG 200
yheO_Trc_E_La.trimmed.seq   --------------------------------------------------


G7715                       CATTGCTAACGGTGAACATACAGGCCGGAAGATTGGTTCGCCAATTACTG 250
yheO_Trc_E_La.trimmed.seq   ----------------CATACAGGCAGGAAGATTGGTTAGCCAACTACTG 34
                                            ********* ************ ***** *****

G7715                       ACCTGGCGCTACGTATGCTGCACGATATGACGGGAGCGGATAGCAGCGTT 300
yheO_Trc_E_La.trimmed.seq   ACCTGGCGCTACGTATGTTGCACGATATGACGGGAGCGGATAGCAGCGTT 84
                            ***************** ********************************

G7715                       TCTAAATGCTACTTTACTCGCGCCAAAAGCGGCGTATTAATGAAGTCCCT 350
yheO_Trc_E_La.trimmed.seq   TTTAAATGCTACATTACTCGCGCCAAAAGCGGCGTATTAATGGAGTCCCT 134
                            * ********** ***************************** *******

G7715                       GACTATCGCGATTCGTAACCGCGAACAGCGTGTAATTGGTCTGCTGTGCA 400
yheO_Trc_E_La.trimmed.seq   GACTATCGCGATTGGTAACCGCGAACAGCGTGTAATTGGTCTGCTGTGCA 184
                            ************ ************************************

G7715                       TCAATATGAATCTTGATGTTCCCTTCTCGCAGATTATGAGCACCTTTGTG 450
yheO_Trc_E_La.trimmed.seq   TCAATATGAATCTTGATGTTCCCTTCTCGCAGATTATGAGCACCTTTGTG 234
                            **************************************************

G7715                       CCGCCAGAAACCCCGGATGTCGGTTCAAGCGTCAACTTTGCCTCTTCTGT 500
yheO_Trc_E_La.trimmed.seq   CCGCCAGAAACCCCGGATGTCGGTTCAAGCGTCAACTTTGCCTGTTCTGT 284
                            ****************************************** ******

G7715                       TGAAGATCTGGTTACCCAAACGCTGGAGTTCACCATCGAAGAAGTGAATG 550
yheO_Trc_E_La.trimmed.seq   CGAAGATCTGGTTACCCAAACGCTGGAGTTCACCATCGAAGAAGTGAATG 334
                             **************************************************

G7715                       CCGATCGCAATGTTTCTAATAACGCCAAAAATCGTCAGATCGTGCTGAAT 600
yheO_Trc_E_La.trimmed.seq   CCGATCGCAATGTTTGTAATAACGCCAAAAATCGTCAGATCGTGCTGAAT 384
                            *************** ********************************

G7715                       CTCTACGAGAAAGGGATCTTCGATATTAAAGATGCGATCAACCAGGTTGC 650
yheO_Trc_E_La.trimmed.seq   CTCTACGAGAAAGGGATCTTCGATATTAAAGATGCGATCAACCAGGTTGC 434
                            **************************************************

G7715                       TGACCGCCTGAACATCTCCAAACACACTGTCTATCTCTACATCCGCCAGT 700
yheO_Trc_E_La.trimmed.seq   TGACCGCCTGAACATCTCCAAACACACTGTCTATCTCTACATCCGCCAGT 484
                            **************************************************

G7715                       TCAAGAGCGGTGATTTCCAGGGGCAAGATAAGTAA--------------- 735
yheO_Trc_E_La.trimmed.seq   TCAAGAGCGGTGATTTCCAGGGGCAAGATAAGTAAAAGGGTTGATCCTAT 534
                            ***********************************

G7715                       ------
yheO_Trc_E_La.trimmed.seq   CGTCAT 540

CLUSTAL W (1.82) multiple sequence alignment
```

```
BACKWARDS!

yhiF_Trc_La.trimmed.seq      --------------------------------------------------
EG11889                      ATGTTTCTTATAATTACCAGGGATACGATGTTCTTCACCGCGATGAAAAA 50


yhiF_Trc_La.trimmed.seq      --------------------------------------------------
EG11889                      CATTCTGAGTAAAGGTAATGTCGTTCATATACAGAACGAAGAAGAGATCG 100


yhiF_Trc_La.trimmed.seq      --------------------------------------------------
EG11889                      ACGTAATGTTGCATCAGAATGCCTTCGTCATTATTGATACATTAATGAAT 150


yhiF_Trc_La.trimmed.seq      --------------------------------------------------
EG11889                      AATGTATTTCATTCTAATTTTCTCACTCAAATTGAACGATTAAAACCTGT 200


yhiF_Trc_La.trimmed.seq      ----------------------------------ACGCTGCCTGGGGAA-G 16
EG11889                      CCATGTCATTATTTTCTCCCCCTTTAATATTAAACGCTGCCTGGGGAAAG 250
                                                 ************** *

yhiF_Trc_La.trimmed.seq      TGCCGGTGACCTTTGTTCCGCGGACTATCACTATCATTGATTTTGTCGCA 66
EG11889                      TGCCGGTGACCTTTGTTCCGCGGACTATCACTATCATTGATTTTGTCGCA 300
                             **************************************************

yhiF_Trc_La.trimmed.seq      CTCATCAATGGCAGTTACTGCTCTGTGCCTGAAGCGGCTGTGTCACTTTC 116
EG11889                      CTCATCAATGGCAGTTACTGCTCTGTGCCTGAAGCGGCTGTGTCACTTTC 350
                             **************************************************

yhiF_Trc_La.trimmed.seq      GCGCAAGCAACATCAGGTTCTGAGCTGCATTGCGAATCAAATGACAACGG 166
EG11889                      GCGCAAGCAACATCAGGTTCTGAGCTGCATTGCGAATCAAATGACAACGG 400
                             **************************************************

yhiF_Trc_La.trimmed.seq      AAGATATTCTGGAGAAACTGAAAATATCGCTAAAAACGTTCTACTGCCAT 216
EG11889                      AAGATATTCTGGAGAAACTGAAAATATCGCTAAAAACGTTCTACTGCCAT 450
                             **************************************************

yhiF_Trc_La.trimmed.seq      AAACACAATATCATGATGATCCTCAATCTTAAGCGGATCAATGAGCTGGT 266
EG11889                      AAACACAATATCATGATGATCCTCAATCTTAAGCGGATCAATGAGCTGGT 500
                             **************************************************

yhiF_Trc_La.trimmed.seq      ACGCCATCAGCATATTGATTATCTGGTGTGAAAGGG 302
EG11889                      ACGCCATCAGCATATTGATTATCTGGTGTGA----- 531
                             *******************************
```

Brief Conclusions: Nine of the twenty sequences made it through the first round of Agencourt sequencing. Seems pretty low, perhaps because I'm not good at the 96-well miniprep right and I didn't have the vacuum regulator yet. Two vectors were ok: pdhR $H\alpha$ and hyaC $L\alpha$. Two in nine is pretty crappy, but much better than one in twenty or so before. Most of the problems are due to fragments being in backwards. There are very few PCR errors this time, so at least one problem is solved. Now I need a way to screen for those reverse guys or to sequence more plasmids per gene.

The beginning sequences of the two correct vectors appeared to be screwed up, containing a deletion each in the 5' upstream TOPO section (not the part I PCR'd). This seemed odd, so I checked it out for a lot of sequences and they all have similar problems (see alignment below). When I looked in more detail at the chromatagraphs however (see Figure 1.20) it was clear that the sequence was ok, it was just a bad phred call.

If I had this to do over, I'd sequence 2-3 of each. Now that I know the pfu Taq improved the error rate.

By gene summary:

- bolA: sequence was complete crap
- cbl: one mutation correct orientation perhaps mutation is sequencing error, but doesn't appear to be
- fecI: both backwards but mutation free

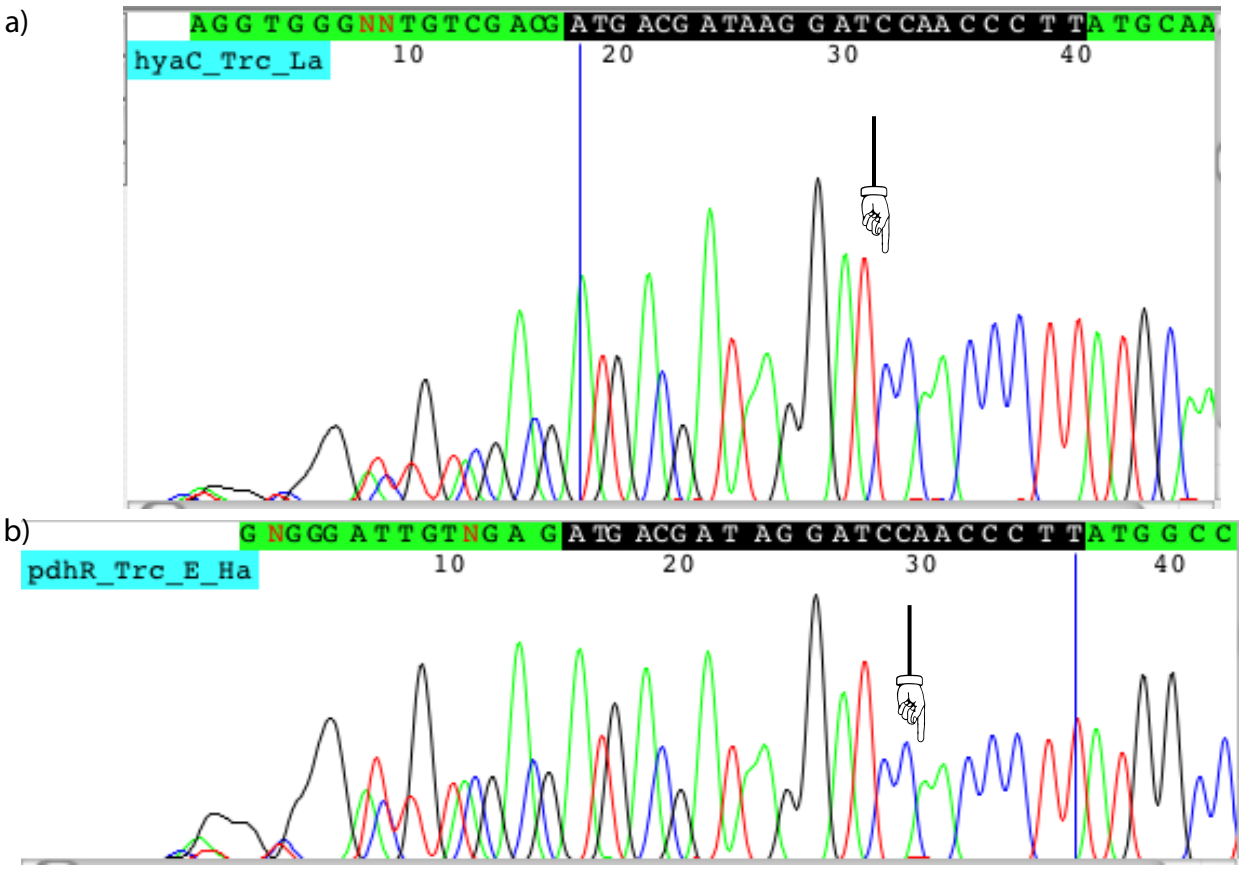


Figure 1.20: hello

- fliA: both backwards, both had mutations (in different spots)
- flhC: backwards
- gadX: both backwards, one had mutation
- glcC: sequence was rubbish
- gfpmutB: both backwards, one had mutation
- *hyaC*: no errors (this is strain La in well C1)
- pdhR: one backwards, one had single bp mutation :(
- *rhaR*: looks good, just the sequence read was too short to tell the end of the sequence need to resequence from other end (this is strain La)
- *sfsA*: backwards with mutation
- *ydaK*: backwards no mutations
- *yheO*: backwards lots of mutations
- *yhiF*: backwards

```
fecI            --CTCTCATCCGCCAAAACAGCCAAGCTTCGAATTCGCCCTT- 40
gfp             TTCTCTCATCCGCCAAAACAGCCAAGCTTCGAATTCGCCCTT- 42
fliA1           ----------------------CAAGCTTCGAATTCGCCCTT- 20
fliA            ----------------------AAGCTTCGAATTCGCCCTT- 19
lrp             -------------------------------------CCCTT- 5
topo            -------GATCTGTACGACGATGACGATAAGGATCCAACCCTT 36
hyaC            --------------------ATGACGATAGGA--TCAACCCTT 21
pdhR            --------------------TGACGATAAGGATCCACCCTT- 21
```

More Sequencing

Tue Nov 15 12:17:21 EST 2005

The low success rate from previous runs and the time-constraint to get this paper out means I gotta focus my effort on fewer TFs. I'm going to pick 8 colonies for the 4 most relevant genes from regulon given our current dataset (lexA, pdhR, fecI, fliA). LexA because we have so many chips run with antibiotics; fliA because the flagellar network has a huge response in many chips; pdhR because of Josh's experimental work on its interaction with the fec genes; fecI because all the algorithms find it as the main iron regulator. In addition two other genes outside regulon are being focused on: abgR (i.e. ydaK) and nusA.

From these eight colonies all found to have inserts will be sent for sequencing (up to five total). Let's hope this works. I'm using the Qiaprep 96-well miniprep again. With the new vacuum, I hope for better results...

Forty-four colonies were picked for the following genes: fecI, ydaK, lexA, nusA, fliA, pdhR, gfp (2-5 colonies per gene). Forty-two colonies grew. Twenty-four were prepped with a Qiagen centrifuge miniprep. The remaining were done in a 96-well plate, the plate contains the following:

96-well primer plate for ChIP-PCR												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	lexA H4	lexA H5	y1	pdhR 1	nus3	pdh4	gfp2	pdh6	fli3	gfp3	gfp5	nus2
B	gfp6	gfp4	pdh2	ydaK1	pdh7	gfp1	-	-	-	-	-	-
C	-	-	-	-	-	-	-	-	-	-	-	-
D	-	-	-	-	-	-	-	-	-	-	-	-
E	-	-	-	-	-	-	-	-	-	-	-	-
F	-	-	-	-	-	-	-	-	-	-	-	-
G	-	-	-	-	-	-	-	-	-	-	-	-
H	-	-	-	-	-	-	-	-	-	-	-	-

Since the agencourt courier arrived faster than I could run my digest on a gel, I didn't get to select genes with an insert. I grabbed twenty-eight sequences and hoped for the best. In addition rhaR La from earlier was cloned into MG1655 and miniprepped. I included it for sequencing with a forward and reverse primer, so we can see if this long sequence is error-free.

The following sequences were sent (big thanks to Ilaria or this would never have been completed on time).

Sample	DNA (ng/μl)	260/280	260/230
D4 rhaR1 (F)	81.7	1.9	2.13
D5 rhaR1 (R)	81.7	1.9	2.13
D6 fec1 (F)	238.4	1.89	2.17
D7 fec2 (F)	57.4	-	-
D8 fec3 (F)	214.6	1.89	2.25
D9 lexA1 (R)	87.4	1.87	2.13
D10 lexA2 (R)	94.6	1.9	1.95
D11 lexA3 (R)	46.9	1.87	2.13
D12 lexA4 (R)	120.3	1.89	1.74
E1 lexA5 (R)	89.1	1.83	1.36
E2 fliA1 (R)	96.5	-	-
E3 fliA2 (R)	63.4	1.85	1.96
E4 fliA3 (R)	71.3	1.85	1.40
E5 fliA4 (R)	88.6	1.84	2.13
E6 fliA5 (R)	71.9	2.01	1.79
E7 pdhR1 (R)	76.0	1.76	1.21
E8 pdhR2 (R)	59.1	1.87	1.77
E9 pdhR4 (R)	66.1	1.89	1.38
E10 pdhR5 (R)	104.3	1.86	2.22
E11 pdhR7 (R)	41.0	1.87	1.77
E12 gfp1 (R)	88.3	1.76	1.40
F1 gfp2 (R)	80.2	1.75	1.27
F2 gfp3 (R)	81.7	1.84	1.64
F3 gfp4 (R)	76.1	1.86	1.66
F4 ydaK1 (R)	59.9	1.80	1.36
F5 ydaK2 (R)	117.3	1.87	2.02
F6 ydaK3 (R)	117.3	1.87	2.02

R = Reverse primer
F = forward primer

The digest (that I didn't use to pick the vectors for sequencing) can be seen in Figure 1.21.

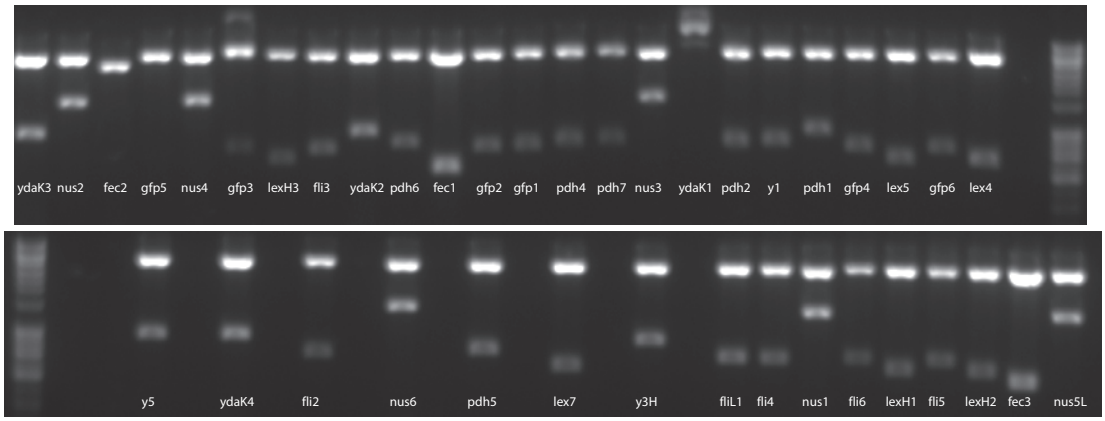


Figure 1.21: 300 ml, 1.5% agarose gel with 4.5 ul of 1% ethidium bromide run for 40 min at 100 volts. 10 μl of Fisher BioReagents exACTGene 1KB Plus DNA Ladder, with bands:weights(ng) of 10000:43, 8000:31, 6000:28, 5000:41, 4000:36, 3000:31, 2500:26, 2000:23, 1500:22, 1000:35, 900:26, 800:52, 700:31, 600:18, 500:35, 400:12, 300:17, 200:12, 100:7 was used. 4.5μl of the centrifuge minipreps was used. 6.0μl of the 96-well plate minipreps was used. y=ydaK, nus=nusA, lex=lexA, fec=fecI

```
CLUSTAL W (1.82) multiple sequence alignment

fecI3 is good!!!!!


fecI3_f.trimmed.seq      ----------CCCTTATGTCTGACCGCGCCACTACCACAGCTTCCTTAACGTTCGAGTCG 50
EG10291                  ----------ATGTCTGACCGCGCCACTACCACAGCTTCCTTAACGTTCGAGTCG 45
```

```
fecI1_f.trimmed.seq      TAGGATCCAACCCTTATGTCTGACCGCGCCACTACCACAGCTTCCTTAACGTTCGAGTCG 60
                                 ******************************************

fecI3_f.trimmed.seq      CTTTATGGCACACATCACGGCTGGTTGAAAAGCTGGCTGACGCGCAAACTCCAGTCTGCT 110
EG10291                  CTTTATGGCACACATCACGGCTGGTTGAAAAGCTGGCTGACGCGCAAACTCCAGTCTGCT 105
fecI1_f.trimmed.seq      CTTTATGGCACACATCACGGCTGGTTGAAAAGCTGGCTGACGCGCAAACTCCAGTCTGCT 120
                         ************************************************************

fecI3_f.trimmed.seq      TTTGATGCAGATGACATTGCCCAGGACACTTTTTTGCGGGTAATGGTCAGCGAAACGCTC 170
EG10291                  TTTGATGCAGATGACATTGCCCAGGACACTTTTTTGCGGGTAATGGTCAGCGAAACGCTC 165
fecI1_f.trimmed.seq      TTTGATGCAGATGACATTGCCCAGGACACTTTTTTGCGGGTAATGGTCAGCGAAACGCTC 180
                         ************************************************************

fecI3_f.trimmed.seq      TCGACGATCCGCGATCCTCGCTCCTTCCTCTGCACTATCGCCAAACGCGTGATGGTGGAC 230
EG10291                  TCGACGATCCGCGATCCTCGCTCCTTCCTCTGCACTATCGCCAAACGCGTGATGGTGGAC 225
fecI1_f.trimmed.seq      TCGACGATCCGCGATCCTCGCTCCTTCCTCTGCACTATCGCCAAACGCGTGATGGTGGAC 240
                         ************************************************************

fecI3_f.trimmed.seq      CTGTTTCGCCGAAACGCGCTGGAAAAAGCGTATCTGGAGATGCTGGCGCTTATGCCGGAG 290
EG10291                  CTGTTTCGCCGAAACGCGCTGGAAAAAGCGTATCTGGAGATGCTGGCGCTTATGCCGGAG 285
fecI1_f.trimmed.seq      CTGTTTCGCCGAAACGCGCTGGAAAAAGCGTATCTGGAGATGCTGGCGCTTATGCCGGAG 300
                         ************************************************************

fecI3_f.trimmed.seq      GGGGGAGCGCCTTCACCTGAGGAACGCGAAAGCCAACTCGAGACCCTACAACTCCTCGAC 350
EG10291                  GGGGGAGCGCCTTCACCTGAGGAACGCGAAAGCCAACTCGAGACCCTACAACTCCTCGAC 345
fecI1_f.trimmed.seq      GGGGGAGCGCCTTCACCTGAGGAACGCGAAAGCCAACTCGAGACCCTACAACTCCTCGAC 360
                         ************************************************************

fecI3_f.trimmed.seq      AGCATGCTGGACGGGCTAAACGGCAAAACACGTGAAGCGTTTCTGCTTTCGCAACTGGAT 410
EG10291                  AGCATGCTGGACGGGCTAAACGGCAAAACACGTGAAGCGTTTCTGCTTTCGCAACTGGAT 405
fecI1_f.trimmed.seq      AGCATGCTGGACGGGCTAAACGGCAAAACACGTGAAGCGTTTCTGCTTTCGCAACTGGAT 420
                         ************************************************************

fecI3_f.trimmed.seq      GGTCTGACATACAGCGAGATTGCGCACAAACTCGGTGTTTCCATCAGCTCCGTGAAAAAA 470
EG10291                  GGTCTGACATACAGCGAGATTGCGCACAAACTCGGTGTTTCCATCAGCTCCGTGAAAAAA 465
fecI1_f.trimmed.seq      GGTCTGACATACAGCGAGATTGCGCACAAACTCGGTGTTTCCATCAGCTCCGTGAAAAAA 480
                         ************************************************************

fecI3_f.trimmed.seq      TACGTGGCGAAAGCCGTCGAGCACTGCCTGCTGTTCCGTCTGGAGTATGGGTTATGAAAG 530
EG10291                  TACGTGGCGAAAGCCGTCGAGCACTGCCTGCTGTTCCGTCTGGAGTATGGGTTATGA--- 522
fecI1_f.trimmed.seq      TACGTGGCGAAAGCCGTCGAGCACTGCCTGCTGTTCCGTCTGGAGTATGGGTTATAA--G 538
                         ********************************************************* *

fecI3_f.trimmed.seq      GGCGAATTCGAAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGAT 590
EG10291                  ------------------------------------------------------------
fecI1_f.trimmed.seq      GGCGAATTCGAAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGAT 598


fecI3_f.trimmed.seq      TAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGT 650
EG10291                  ------------------------------------------------------------
fecI1_f.trimmed.seq      TAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGT 658


fecI3_f.trimmed.seq      GGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGT 710
EG10291                  ------------------------------------------------------------
fecI1_f.trimmed.seq      GGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGT 718


fecI3_f.trimmed.seq      GGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGT 770
EG10291                  ------------------------------------------------------------
fecI1_f.trimmed.seq      GGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGT 778


fecI3_f.trimmed.seq      CGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGA 830
EG10291                  ------------------------------------------------------------
fecI1_f.trimmed.seq      CGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGA 838


fecI3_f.trimmed.seq      CAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAC 868
EG10291                  ---------------------------------------
fecI1_f.trimmed.seq      CAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAG--- 873


CLUSTAL W (1.82) multiple sequence alignment

fliA2 and fliA5 are GOOD!
all others are backwards


fliA1_r.trimmed.seq      ------------------------------------------------------------
fliA4_r.trimmed.seq      ------------------------------------------------------------
fliA3_r.trimmed.seq      ------------------------------------------------------------
EG11355                  ------------------------------------------------------------
fliA2_r.trimmed.seq      TTAACAATTTATCAGACAATCTGTGTGGCCACTCGACCGGAATTATCGATTAACTTTATT 60
fliA5_r.trimmed.seq      TTAACA-TTTATCAGACAATCTGTGTGGGCACTCGACCGGAATTATCGATTAACTTTATT 59


fliA1_r.trimmed.seq      ------------------------------------------------------------
fliA4_r.trimmed.seq      ------------------------------------------------------------
fliA3_r.trimmed.seq      ------------------------------------------------------------
EG11355                  ------------------------------------------------------------
fliA2_r.trimmed.seq      ATTAAAAATTAAAGAGGTATATATTAATGTATCGATTAAATAAGGAGGAATAAACCATGG 120
```

```
fliA5_r.trimmed.seq      ATTAAAAATTAAAGAGGTATATATTAATGTATCGATTAAATAAGGAGGAATAAACCATGG 119


fliA1_r.trimmed.seq      ------------------------------------------------------------
fliA4_r.trimmed.seq      ------------------------------------------------------------
fliA3_r.trimmed.seq      ------------------------------------------------------------
EG11355                  ------------------------------------------------------------
fliA2_r.trimmed.seq      GGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG 180
fliA5_r.trimmed.seq      GGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG 179


fliA1_r.trimmed.seq      --------------CGCAACGCCA-GCTTCGATTCGCCCTTGTGAATTCACTCTATACCG 45
fliA4_r.trimmed.seq      --------------------------------TTCGCCCTTGTGAATTCACTCTATACCG 28
fliA3_r.trimmed.seq      ----------TTTTCGCAACGCCAAGCTTCGATTCGCCCTTGTGAATTCACTCTATACCG 50
EG11355                  -------------------------------------------GTGAATTCACTCTATACCG 19
fliA2_r.trimmed.seq      GTCGGGATCTGTACGACGATGACGATAAGGATCCAACCCTTGTGAATTCACTCTATACCG 240
fliA5_r.trimmed.seq      GTCGGGATCTGTACGACGATGACGATAAGGATCCAACCCTTGTGAATTCACTCTATACCG 239
                                                           ******************


fliA1_r.trimmed.seq      CTGAAGGTGTAATGGATAAACACTCGCTGTGGCAGCGTTATGTCCCGCTGGTGCGTCACG 105
fliA4_r.trimmed.seq      CTGAAGGTGTAATGGATAAACACTCGCTGTGGCAGCGTTATGTCCCGCTGGTGCGTCACG 88
fliA3_r.trimmed.seq      CTGAAGGTGTAATGGATAAACACTCGCTGTGGCAGCGTTATGTCCCGCTGGTGCGTCACG 110
EG11355                  CTGAAGGTGTAATGGATAAACACTCGCTGTGGCAGCGTTATGTCCCGCTGGTGCGTCACG 79
fliA2_r.trimmed.seq      CTGAAGGTGTAATGGATAAACACTCGCTGTGGCAGCGTTATGTCCCGCTGGTGCGTCACG 300
fliA5_r.trimmed.seq      CTGAAGGTGTAATGGATAAACACTCGCTGTGGCAGCGTTATGTCCCGCTGGTGCGTCACG 299
                         ************************************************************


fliA1_r.trimmed.seq      AAGCATTGCGCCTGCAGGTTCGACTGCCCGCGAGCGTGGAACTTGACGATCTGCTACAGG 165
fliA4_r.trimmed.seq      AAGCATTGTGCCTGCAGGTTCGACTGCCCGCGAGCGTGGAACTTGACGATCTGCTACAGG 148
fliA3_r.trimmed.seq      AAGCATTGCGCCTGCAGGTTCGACTGCCCGCGAGCGTGGAACTTGACGATCTGCTACAGG 170
EG11355                  AAGCATTGCGCCTGCAGGTTCGACTGCCCGCGAGCGTGGAACTTGACGATCTGCTACAGG 139
fliA2_r.trimmed.seq      AAGCATTGCGCCTGCAGGTTCGACTGCCCGCGAGCGTGGAACTTGACGATCTGCTACAGG 360
fliA5_r.trimmed.seq      AAGCATTGCGCCTGCAGGTTCGACTGCCCGCGAGCGTGGAACTTGACGATCTGCTACAGG 359
                         ******** **************************************************


fliA1_r.trimmed.seq      CGGGCGGCATTGGGTTACTTAATGCCGTCGAACGCTATGACGCCCTACAAGGAACGGCAT 225
fliA4_r.trimmed.seq      CGGGCGGCATTGGGTTACTTAATGCCGTCGAACGCTATGACGCCCTACAAGGAACGGCAT 208
fliA3_r.trimmed.seq      CGGGCGGCATTGGGTTACTTAATGCCGTCGAACGCTATGACGCCCTACAAGGAACGGCAT 230
EG11355                  CGGGCGGCATTGGGTTACTTAATGCCGTCGAACGCTATGACGCCCTACAAGGAACGGCAT 199
fliA2_r.trimmed.seq      CGGGCGGCATTGGGTTACTTAATGCCGTCGAACGCTATGACGCCCTACAAGGAACGGCAT 420
fliA5_r.trimmed.seq      CGGGCGGCATTGGGTTACTTAATGCCGTCGAACGCTATGACGCCCTACAAGGAACGGCAT 419
                         ************************************************************


fliA1_r.trimmed.seq      TTACAACTTACGCAGTGCAGCGTATCCGTGGCGCTATGCTGGATGAACTTCGCAGCCGTG 285
fliA4_r.trimmed.seq      TTACAACTTACGCAGTGCAGCGTATCCGTGGCGCTATGCTGGATGAACTTCGCAGCCGTG 268
fliA3_r.trimmed.seq      TTACAACTTACGCAGTGCAGCGTATCCGTGGCGCTATGCTGGATGAACTTCGCAGCCGTG 290
EG11355                  TTACAACTTACGCAGTGCAGCGTATCCGTGGCGCTATGCTGGATGAACTTCGCAGCCGTG 259
fliA2_r.trimmed.seq      TTACAACTTACGCAGTGCAGCGTATCCGTGGCGCTATGCTGGATGAACTTCGCAGCCGTG 480
fliA5_r.trimmed.seq      TTACAACTTACGCAGTGCAGCGTATCCGTGGCGCTATGCTGGATGAACTTCGCAGCCGTG 479
                         ************************************************************


fliA1_r.trimmed.seq      ACTGGGTGCCGCGCAGCGTGCGACGCAACGCGCGTGAAGTGGCACAGGCAATAGGGCAAC 345
fliA4_r.trimmed.seq      ACTGGGTGCCGCGCAGCGTGCGACGCAACGCGCGTGAAGTGGCACAGGCAATAGGGCAAC 328
fliA3_r.trimmed.seq      ACTGGGTGCCGCGCAGCGTGCGACGCAACGCGCGTGAAGTGGCACAGGCAATAGGGCAAC 350
EG11355                  ACTGGGTGCCGCGCAGCGTGCGACGCAACGCGCGTGAAGTGGCACAGGCAATAGGGCAAC 319
fliA2_r.trimmed.seq      ACTGGGTGCCGCGCAGCGTGCGACGCAACGCGCGTGAAGTGGCACAGGCAATAGGGCAAC 540
fliA5_r.trimmed.seq      ACTGGGTGCCGCGCAGCGTGCGACGCAACGCGCGTGAAGTGGCACAGGCAATAGGGCAAC 539
                         ************************************************************


fliA1_r.trimmed.seq      TGGAGCAGGAACTTGGCCGCAACGCCACGGAAACTGAGGTAGCGGAACGTTTAGGGATCG 405
fliA4_r.trimmed.seq      TGGAGCAGGAACTTGGCCGCAACGCCACGGAAACTGAGGTAGCGGAACGTTTAGGGATCG 388
fliA3_r.trimmed.seq      TGGAGCAGGAACTTGGCCGCAACGCCACGGAAACTGAGGTAGCGGAACGTTTAGGGATCG 410
EG11355                  TGGAGCAGGAACTTGGCCGCAACGCCACGGAAACTGAGGTAGCGGAACGTTTAGGGATCG 379
fliA2_r.trimmed.seq      TGGAGCAGGAACTTGGCCGCAACGCCACGGAAACTGAGGTAGCGGAACGTTTAGGGATCG 600
fliA5_r.trimmed.seq      TGGAGCAGGAACTTGGCCGCAACGCCACGGAAACTGAGGTAGCGGAACGTTTAGGGATCG 599
                         ************************************************************


fliA1_r.trimmed.seq      ATATTGCCGATTATCGCCAAATGTTGCTCGACACCAATAACAGCCAGCTCTTCTCCTACG 465
fliA4_r.trimmed.seq      ATATTGCCGATTATCGCCAAATGTTGCTCGACACCAATAACAGCCAGCTCTTCTCCTACG 448
fliA3_r.trimmed.seq      ATATTGCCGATTATCGCCAAATGTTGCTCGACACCAATAACAGCCAGCTCTTCTCCTACG 470
EG11355                  ATATTGCCGATTATCGCCAAATGTTGCTCGACACCAATAACAGCCAGCTCTTCTCCTACG 439
fliA2_r.trimmed.seq      ATATTGCCGATTATCGCCAAATGTTGCTCGACACCAATAACAGCCAGCTCTTCTCCTACG 660
fliA5_r.trimmed.seq      ATATTGCCGATTATCGCCAAATGTTGCTCGACACCAATAACAGCCAGCTCTTCTCCTACG 659
                         ************************************************************


fliA1_r.trimmed.seq      ATGAGTGGCGCGAAGAGCACGGCGATAGCATCGAACTGGTTACTGATGATCATCAGCGAG 525
fliA4_r.trimmed.seq      ATGAGTGGCGCGAAGAGCACGGCGATAGCATCGAACTGGTTACTGATGATCATCAGCGAG 508
fliA3_r.trimmed.seq      ATGAGTGGCGCGAAGAGCACGGCGATAGCATCGAACTGGTTACTGATGATCATCAGCGAG 530
EG11355                  ATGAGTGGCGCGAAGAGCACGGCGATAGCATCGAACTGGTTACTGATGATCATCAGCGAG 499
fliA2_r.trimmed.seq      ATGAGTGGCGCGAAGAGCACGGCGATAGCATCGAACTGGTTACTGATGATCATCAGCGAG 720
fliA5_r.trimmed.seq      ATGAGTGGCGCGAAGAGCACGGCGATAGCATCGAACTGGTTACTGATGATCATCAGCGAG 719
                         ************************************************************


fliA1_r.trimmed.seq      AAAACCCGCTACAACAACTACTGGACAGTAATCTGCGCCAGCGGGTGATGGAAGCCATCG 585
fliA4_r.trimmed.seq      AAAACCCGCTACAACAACTACTGGACAGTAATCTGCGCCAGCGGGTGATGGAAGCCATCG 568
fliA3_r.trimmed.seq      AAAACCCGCTACAACAACTACTGGACAGTAATCTGCGCCAGCGGGTGATGGAAGCCATCG 590
EG11355                  AAAACCCGCTACAACAACTACTGGACAGTAATCTGCGCCAGCGGGTGATGGAAGCCATCG 559
fliA2_r.trimmed.seq      AAAACCCGCTACAACAACTACTGGACAGTAATCTGCGCCAGCGGGTGATGGAAGCCATCG 780
fliA5_r.trimmed.seq      AAAACCCGCTACAACAACTACTGGACAGTAATCTGCGCCAGCGGGTGATGGAAGCCATCG 779
                         ************************************************************


fliA1_r.trimmed.seq      AAACGTTGCCGGAGCGCGAAAAACTGGTATTAACCCTCTATTACCAGGAAGAGCTGAATC 645
```

```
fliA4_r.trimmed.seq      AAACGTTGCCGGAGCGCGAAAAACTGGTATTAACCCTCTATTACCAGGAAGAGCTGAATC 628
fliA3_r.trimmed.seq      AAACGTTGCCGGAGCGCGAAAAACTGGTATTAACCCTCTATTACCAGGAAGAGCTGAATC 650
EG11355                  AAACGTTGCCGGAGCGCGAAAAACTGGTATTAACCCTCTATTACCAGGAAGAGCTGAATC 619
fliA2_r.trimmed.seq      AAACGTTGCCGGAGCGCGAAAAACTGGTATTAACCCTCTATTACCAGGAAGAGCTGAATC 840
fliA5_r.trimmed.seq      AAACGTTGCCGGAGCGCGAAAAACTGGTATTAACCCTCTATTACCAGGAAGAGCTGAATC 839
                         ************************************************************

fliA1_r.trimmed.seq      TCAAAGAGATTGGCGCGGTGCTGGAGGTCGGGGAATCGCGGGTCAGTCAGTTACACAGCC 705
fliA4_r.trimmed.seq      TCAAAGAGATTGGCGCGGTGCTGGAGGTCGGGGAATCGCGGGTCAGTCAGTTACACAGCC 688
fliA3_r.trimmed.seq      TCAAAGAGATTGGCGCGGTGCTGGAGGTCGGGGAATCGCGGGTCAGTCAGTTACACAGCC 710
EG11355                  TCAAAGAGATTGGCGCGGTGCTGGAGGTCGGGGAATCGCGGGTCAGTCAGTTACACAGCC 679
fliA2_r.trimmed.seq      TCAAAGAGATTGGCGCGGTGCTGGAGGTCGGGGAATCGCGGGTCAGTCAGTTACACAGCC 900
fliA5_r.trimmed.seq      TCAAAGAGATTGGCGCGGTGCTGGAGGTCGGGGAATCGCGGGTCAGTCAGTTACACAGCC 899
                         ************************************************************

fliA1_r.trimmed.seq      AGGCTATTAAACGGTTACGCACTAAACTGGGTAAGTTATAAAAGGGTTGGATCCTTATCG 765
fliA4_r.trimmed.seq      AGGCTATTAAACGGTTACGCACTAAACTGGGTAAGTTATAAAAGGGTTGGATCCTTATCG 748
fliA3_r.trimmed.seq      AGGCTATTAAACGGTTACGCACTAAACTGGGTAAGTTATAAAAGGGTTGGATCCTTATCG 770
EG11355                  AGGCTATTAAACGGTTACGCACTAAACTGGGTAAGTTATAA-------------------- 720
fliA2_r.trimmed.seq      AGGCTATTAAACGGTTACGCACTAAACTGGGTAAGTT-TAAAAGGGC-GAAT-CGAA--G 955
fliA5_r.trimmed.seq      AGGCTATTAAACGGTTACGCACTAAACTGGGTAAGTT-TAAAAGGGC-GAATTCGAA--G 955
                         ************************************ ***

fliA1_r.trimmed.seq      TCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACCATGA 825
fliA4_r.trimmed.seq      TCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACCATGA 808
fliA3_r.trimmed.seq      TCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACCATGA 830
EG11355                  ------------------------------------------------------------
fliA2_r.trimmed.seq      CTGGCGTTGCG---------------------------------------------------- 966
fliA5_r.trimmed.seq      CTGGCGTTGCG---------------------------------------------------- 966


fliA1_r.trimmed.seq      TGATGATGATGATGAGAACCCCCCATGGTTTATTCCTCCTTATTTAATCGATACATTAAT 885
fliA4_r.trimmed.seq      TGATGATGATGATGAGAACCCCCCATGGTTTATTCCTCCTTATTTAATCGATACATTAAT 868
fliA3_r.trimmed.seq      TGATGATGATGATGAGAACCCCCCATGGTTTATTCCTCCTTATTTAATCGATACATTAAT 890
EG11355                  ------------------------------------------------------------
fliA2_r.trimmed.seq      ------------------------------------------------------------
fliA5_r.trimmed.seq      ------------------------------------------------------------


fliA1_r.trimmed.seq      ATATACCTCTTTAATTTTTAATAATAAAGTTAATCGATAATTCCGGTCGAGTGCCCACAC 945
fliA4_r.trimmed.seq      ATATACCTCTTTTATTTTTAATAATAAAGTTAATCGATAATTCCGGTCGAGTGCCCACAC 928
fliA3_r.trimmed.seq      ATATACCTCTTTAATTTTTAATAATAAAGT-AATCGATAATTCCGGT------------- 936
EG11355                  ------------------------------------------------------------
fliA2_r.trimmed.seq      ------------------------------------------------------------
fliA5_r.trimmed.seq      ------------------------------------------------------------


fliA1_r.trimmed.seq      AGATTGTCTGATAAATTG 963
fliA4_r.trimmed.seq      AGATTGTCTGATAAAT-- 944
fliA3_r.trimmed.seq      ------------------
EG11355                  ------------------
fliA2_r.trimmed.seq      ------------------
fliA5_r.trimmed.seq      ------------------



CLUSTAL W (1.82) multiple sequence alignment

gfp1 is backwards; all others are forwards.
gfp3 might have mutation.
gfp2 and gfp4 both look good.  I'd keep gfp4


gfp3_r.trimmed.seq       ----------------------------------------AATTATCGATTAACTTTATTA 21
gfp4_r.trimmed.seq       TTAACATTTATCAGACAATCTGTGTGGGCACTCGACCGGAATTATCGATTAACTTTATTA 60
gfp2_r.trimmed.seq       ------------------------------------------------------TATTA 5
Gfpmut3b                 ------------------------------------------------------------
gfp1_r.trimmed.seq       ------------------------------------------------------------


gfp3_r.trimmed.seq       TTAAAAATTAAAGAGGTATATATTAATGTATCGATTAAATAAGGAGGAATAAACCATGGG 81
gfp4_r.trimmed.seq       TTAAAAATTAAAGAGGTATATATTAATGTATCGATTAAATAAGGAGGAATAAACCATGGG 120
gfp2_r.trimmed.seq       TTAAAAATTAAAGAGGTATATATTAATGTATCGATTAAATAAGGAGGAATAAACCATGGG 65
Gfpmut3b                 ------------------------------------------------------------
gfp1_r.trimmed.seq       ------------------------------------------------------------


gfp3_r.trimmed.seq       GGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGG 141
gfp4_r.trimmed.seq       GGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGG 180
gfp2_r.trimmed.seq       GGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGG 125
Gfpmut3b                 ------------------------------------------------------------
gfp1_r.trimmed.seq       ------------------------------------------------------------


gfp3_r.trimmed.seq       TCGGGATCTGTACGACGATGACGATAAGGATCCAACCCTTATGCGTAAAGGAGAAGAACT 201
gfp4_r.trimmed.seq       TCGGGATCTGTACGACGATGACGATAAGGATCCAACCCTTATGCGTAAAGGAGAAGAACT 240
gfp2_r.trimmed.seq       TCGGGATCTGTACGACGATGACGATAAGGATCCAACCCTTATGCGTAAAGGAGAAGAACT 185
Gfpmut3b                 ----------------------------------------ATGCGTAAAGGAGAAGAACT 20
gfp1_r.trimmed.seq       --------TTTTCG-CAACGCCAGCTTCGATTCG-CCCTTATGCGTAAAGGAGAAGAACT 50
                                                                 ********************

gfp3_r.trimmed.seq       TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATT 261
```

```
gfp4_r.trimmed.seq      TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATT 300
gfp2_r.trimmed.seq      TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATT 245
Gfpmut3b                TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATT 80
gfp1_r.trimmed.seq      TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATT 110
                        ************************************************************

gfp3_r.trimmed.seq      TTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTAT 321
gfp4_r.trimmed.seq      TTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTAT 360
gfp2_r.trimmed.seq      TTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTAT 305
Gfpmut3b                TTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTAT 140
gfp1_r.trimmed.seq      TTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTAT 170
                        ************************************************************

gfp3_r.trimmed.seq      TTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGTTATGG 381
gfp4_r.trimmed.seq      TTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGTTATGG 420
gfp2_r.trimmed.seq      TTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGTTATGG 365
Gfpmut3b                TTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGTTATGG 200
gfp1_r.trimmed.seq      TTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGTTATGG 230
                        ************************************************************

gfp3_r.trimmed.seq      TGTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGC 441
gfp4_r.trimmed.seq      TGTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGC 480
gfp2_r.trimmed.seq      TGTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGC 425
Gfpmut3b                TGTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGC 260
gfp1_r.trimmed.seq      TGTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGC 290
                        ************************************************************

gfp3_r.trimmed.seq      CATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAA 501
gfp4_r.trimmed.seq      CATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAA 540
gfp2_r.trimmed.seq      CATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAA 485
Gfpmut3b                CATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAA 320
gfp1_r.trimmed.seq      CATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAA 350
                        ************************************************************

gfp3_r.trimmed.seq      GACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGG 561
gfp4_r.trimmed.seq      GACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGG 600
gfp2_r.trimmed.seq      GACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGG 545
Gfpmut3b                GACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGG 380
gfp1_r.trimmed.seq      GACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGG 410
                        ************************************************************

gfp3_r.trimmed.seq      TATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTC 621
gfp4_r.trimmed.seq      TATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTC 660
gfp2_r.trimmed.seq      TATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTC 605
Gfpmut3b                TATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTC 440
gfp1_r.trimmed.seq      TATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTC 470
                        ************************************************************

gfp3_r.trimmed.seq      ACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAAT 681
gfp4_r.trimmed.seq      ACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAAT 720
gfp2_r.trimmed.seq      ACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAAT 665
Gfpmut3b                ACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAAT 500
gfp1_r.trimmed.seq      ACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAAT 530
                        ************************************************************

gfp3_r.trimmed.seq      TAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCC 741
gfp4_r.trimmed.seq      TAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCC 780
gfp2_r.trimmed.seq      TAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCC 725
Gfpmut3b                TAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCC 560
gfp1_r.trimmed.seq      TAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCC 590
                        ************************************************************

gfp3_r.trimmed.seq      AATTGGCGATGGCCCCGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCT 801
gfp4_r.trimmed.seq      AATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCT 840
gfp2_r.trimmed.seq      AATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCT 785
Gfpmut3b                AATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCT 620
gfp1_r.trimmed.seq      AATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCT 650
                        *************** ****************************************

gfp3_r.trimmed.seq      TTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGC 861
gfp4_r.trimmed.seq      TTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGC 900
gfp2_r.trimmed.seq      TTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGC 845
Gfpmut3b                TTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGC 680
gfp1_r.trimmed.seq      TTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGC 710
                        ************************************************************

gfp3_r.trimmed.seq      TGGGATTACACATGGCATGGATGAACTATACAAATAAAAGGGC-GAAT-CGAA--GCTGG 917
gfp4_r.trimmed.seq      TGGGATTACACATGGCATGGATGAACTATACAAATAAAAGGGC-GAAT-CGAA--GCTGG 956
gfp2_r.trimmed.seq      TGGGATTACACATGGCATGGATGAACTATACAAATAAAAGGGC-GAATTCGAA--GC--- 899
Gfpmut3b                TGGGATTACACATGGCATGGATGAACTATACAAATAA---------------------- 717
gfp1_r.trimmed.seq      TGGGATTACACATGGCATGGATGAACTATACAAATAAAAGGGTTGGATCCTTATCGTCAT 770
                        ************************************

gfp3_r.trimmed.seq      CGTTGCGAAAAA----------------------------------------------- 929
gfp4_r.trimmed.seq      CGTTGCG---------------------------------------------------- 963
gfp2_r.trimmed.seq      ----------------------------------------------------------
Gfpmut3b                ----------------------------------------------------------
gfp1_r.trimmed.seq      CGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACCATGATGAT 830


gfp3_r.trimmed.seq      ----------------------------------------------------------
gfp4_r.trimmed.seq      ----------------------------------------------------------
```

```
gfp2_r.trimmed.seq      ------------------------------------------------------------
Gfpmut3b                ------------------------------------------------------------
gfp1_r.trimmed.seq      GATGATGATGAGAACCCCCCATGGTTTATTCCTCCTTATTTAATCGATACATTAATATAT 890


gfp3_r.trimmed.seq      ----------------------
gfp4_r.trimmed.seq      ----------------------
gfp2_r.trimmed.seq      ----------------------
Gfpmut3b                ----------------------
gfp1_r.trimmed.seq      ACCTCTTTATTTTTTAATAATAA 913


CLUSTAL W (1.82) multiple sequence alignment

lexA4 is GOOD!!!!!
all others backwards.

lexA2_r.trimmed.seq     ------------------------------------------------------------
lexA5_r.trimmed.seq     ------------------------------------------------------------
lexA1_r.trimmed.seq     ------------------------------------------------------------
lexA3_r.trimmed.seq     ------------------------------------------------------------
EG10533                 ------------------------------------------------------------
lexA4_r.trimmed.seq     GAAATGAGCTGTGACAATTTATTCATCCGGCTCGTATAATGTGTGGAATTGTGAGCGGAT 60


lexA2_r.trimmed.seq     ------------------------------------------------------------
lexA5_r.trimmed.seq     ------------------------------------------------------------
lexA1_r.trimmed.seq     ------------------------------------------------------------
lexA3_r.trimmed.seq     ------------------------------------------------------------
EG10533                 ------------------------------------------------------------
lexA4_r.trimmed.seq     AACAATTTCACACAGGAAACAGCGCCGCTGAGAAAAAGCGAAGCGGCACTGCTCTTTAAC 120


lexA2_r.trimmed.seq     ------------------------------------------------------------
lexA5_r.trimmed.seq     ------------------------------------------------------------
lexA1_r.trimmed.seq     ------------------------------------------------------------
lexA3_r.trimmed.seq     ------------------------------------------------------------
EG10533                 ------------------------------------------------------------
lexA4_r.trimmed.seq     AATTTATCAGACAATCTGTGTGGGCACTCGACCGGAATTATCGATTAACTTTATTATTAA 180


lexA2_r.trimmed.seq     ------------------------------------------------------------
lexA5_r.trimmed.seq     ------------------------------------------------------------
lexA1_r.trimmed.seq     ------------------------------------------------------------
lexA3_r.trimmed.seq     ------------------------------------------------------------
EG10533                 ------------------------------------------------------------
lexA4_r.trimmed.seq     AAATTAAAGAGGTATATATTAATGTATCGATTAAATAAGGAGGAATAAACCATGGGGGGT 240


lexA2_r.trimmed.seq     ------------------------------------------------------------
lexA5_r.trimmed.seq     ------------------------------------------------------------
lexA1_r.trimmed.seq     ------------------------------------------------------------
lexA3_r.trimmed.seq     ------------------------------------------------------------
EG10533                 ------------------------------------------------------------
lexA4_r.trimmed.seq     TCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGG 300


lexA2_r.trimmed.seq     ----------------------------TTCGCCCTTATGAAAGCGTTAACGGCCAGGCAA 33
lexA5_r.trimmed.seq     --------CGCAACAGCCAGCTTCGAATTCGCCCTTATGAAAGCGTTAACGGCCAGGCAA 52
lexA1_r.trimmed.seq     ----------------------------TTCGCCCTTATGAAAGCGTTAACGGCCAGGCAA 33
lexA3_r.trimmed.seq     ----------------CCAAGCTTCGATTCGCCCTTATGAAAGCGTTAACGGCCAGGCAA 44
EG10533                 ----------------------------------ATGAAAGCGTTAACGGCCAGGCAA 24
lexA4_r.trimmed.seq     GATCTGTACGACGATGACGATAAGGATCCAACCCTTATGAAAGCGTTAACGGCCAGGCAA 360
                                                          ************************

lexA2_r.trimmed.seq     CAAGAGGTGTTTGATCTCATCCGTGATCACATCAGCCAGACAGGTATGCCGCCGACGCGT 93
lexA5_r.trimmed.seq     CAAGAGGTGTTTGATCTCATCCGTGATCACATCAGCCAGACAGGTATGCCGCCGACGCGT 112
lexA1_r.trimmed.seq     CAAGAGGTGTTTGATCTCATCCGTGATCACATCAGCCAGACAGGTATGCCGCCGACGCGT 93
lexA3_r.trimmed.seq     CAAGAGGTGTTTGATCTCATCCGTGATCACATCAGCCAGACAGGTATGCCGCCGACGCGT 104
EG10533                 CAAGAGGTGTTTGATCTCATCCGTGATCACATCAGCCAGACAGGTATGCCGCCGACGCGT 84
lexA4_r.trimmed.seq     CAAGAGGTGTTTGATCTCATCCGTGATCACATCAGCCAGACAGGTATGCCGCCGACGCGT 420
                        ************************************************************

lexA2_r.trimmed.seq     GCGGAAATCGCGCAGCGTTTGGGGTTCCGTTCCCCAAACGCGGCTGAAGAACATCTGAAG 153
lexA5_r.trimmed.seq     GCGGAAATCGCGCAGCGTTTGGGGTTCCCCAAACGCGGCTGAAGAACATCTGAAG 172
lexA1_r.trimmed.seq     GCGGAAATCGCGCAGCGTTTGGGGTTCCGTTCCCCAAACGCGGCTGAAGAACATCTGAAG 153
lexA3_r.trimmed.seq     GCGGAAATCGCGCAGCGTTTGGGGTTCCGTTCCCCAAACGCGGCTGAAGAACATCTGAAG 164
EG10533                 GCGGAAATCGCGCAGCGTTTGGGGTTCCGTTCCCCAAACGCGGCTGAAGAACATCTGAAG 144
lexA4_r.trimmed.seq     GCGGAAATCGCGCAGCGTTTGGGGTTCCGTTCCCCAAACGCGGCTGAAGAACATCTGAAG 480
                        ************************************************************

lexA2_r.trimmed.seq     GCGCTGGCACGCAAAGGCGTTATTGAAATTGTTTCCGGCGCATCACGCGGGATTCGTCTG 213
lexA5_r.trimmed.seq     GCGCTGGCACGCAAAGGCGTTATTGAAATTGTTTCCGGCGCATCACGCGGGATTCGTCTG 232
lexA1_r.trimmed.seq     GCGCTGGCACGCAAAGGCGTTATTGAAATTGTTTCCGGCGCATCACGCGGGATTCGTCTG 213
lexA3_r.trimmed.seq     GCGCTGGCACGCAAAGGCGTTATTGAAATTGTTTCCGGCGCATCACGCGGGATTCGTCTG 224
EG10533                 GCGCTGGCACGCAAAGGCGTTATTGAAATTGTTTCCGGCGCATCACGCGGGATTCGTCTG 204
lexA4_r.trimmed.seq     GCGCTGGCACGCAAAGGCGTTATTGAAATTGTTTCCGGCGCATCACGCGGGATTCGTCTG 540
                        ************************************************************

lexA2_r.trimmed.seq     TTGCAGGAAGAGGAAGAAGGGTTGCCGCTGGTAGGTCGTGTGGCTGCCGGTGAACCACTT 273
lexA5_r.trimmed.seq     TTGCAGGAAGAGGAAGAAGGGTTGCCGCTGGTAGGTCGTGTGGCTGCCGGTGAACCACTT 292
lexA1_r.trimmed.seq     TTGCAGGAAGAGGAAGAAGGGTTGCCGCTGGTAGGTCGTGTGGCTGCCGGTGAACCACTT 273
```

```
lexA3_r.trimmed.seq      TTGCAGGAAGAGGAAGAAGGGTTGCCGCTGGTAGGTCGTGTGGCTGCCGGTGAACCACTT 284
EG10533                  TTGCAGGAAGAGGAAGAAGGGTTGCCGCTGGTAGGTCGTGTGGCTGCCGGTGAACCACTT 264
lexA4_r.trimmed.seq      TTGCAGGAAGAGGAAGAAGGGTTGCCGCTGGTAGGTCGTGTGGCTGCCGGTGAACCACTT 600
                         ************************************************************

lexA2_r.trimmed.seq      CTGGCGCAACAGCATATTGAAGGTCATTATCAGGTCGATCCTTCCTTATTCAAGCCGAAT 333
lexA5_r.trimmed.seq      CTGGCGCAACAGCATATTGAAGGTCATTATCAGGTCGATCCTTCCTTATTCAAGCCGAAT 352
lexA1_r.trimmed.seq      CTGGCGCAACAGCATATTGAAGGTCATTATCAGGTCGATCCTTCCTTATTCAAGCCGAAT 333
lexA3_r.trimmed.seq      CTGGCGCAACAGCATATTGAAGGTCATTATCAGGTCGATCCTTCCTTATTCAAGCCGAAT 344
EG10533                  CTGGCGCAACAGCATATTGAAGGTCATTATCAGGTCGATCCTTCCTTATTCAAGCCGAAT 324
lexA4_r.trimmed.seq      CTGGCGCAACAGCATATTGAAGGTCATTATCAGGTCGATCCTTCCTTATTCAAGCCGAAT 660
                         ************************************************************

lexA2_r.trimmed.seq      GCTGATTTCCTGCTGCGCGTCAGCGGGATGTCGATGAAAGATATCGGCATTATGGATGGT 393
lexA5_r.trimmed.seq      GCTGATTTCCTGCTGCGCGTCAGCGGGATGTCGATGAAAGATATCGGCATTATGGATGGT 412
lexA1_r.trimmed.seq      GCTGATTTCCTGCTGCGCGTCAGCGGGATGTCGATGAAAGATATCGGCATTATGGATGGT 393
lexA3_r.trimmed.seq      GCTGATTTCCTGCTGCGCGTCAGCGGGATGTCGATGAAAGATATCGGCATTATGGATGGT 404
EG10533                  GCTGATTTCCTGCTGCGCGTCAGCGGGATGTCGATGAAAGATATCGGCATTATGGATGGT 384
lexA4_r.trimmed.seq      GCTGATTTCCTGCTGCGCGTCAGCGGGATGTCGATGAAAGATATCGGCATTATGGATGGT 720
                         ************************************************************

lexA2_r.trimmed.seq      GACTTGCTGGCAGTGCATAAAACTCAGGATGTACGTAACGGTCAGGTCGTTGTCGCACGT 453
lexA5_r.trimmed.seq      GACTTGCTGGCAGTGCATAAAACTCAGGATGTACGTAACGGTCAGGTCGTTGTCGCACGT 472
lexA1_r.trimmed.seq      GACTTGCTGGCAGTGCATAAAACTCAGGATGTACGTAACGGTCAGGTCGTTGTCGCACGT 453
lexA3_r.trimmed.seq      GACTTGCTGGCAGTGCATAAAACTCAGGATGTACGTAACGGTCAGGTCGTTGTCGCACGT 464
EG10533                  GACTTGCTGGCAGTGCATAAAACTCAGGATGTACGTAACGGTCAGGTCGTTGTCGCACGT 444
lexA4_r.trimmed.seq      GACTTGCTGGCAGTGCATAAAACTCAGGATGTACGTAACGGTCAGGTCGTTGTCGCACGT 780
                         ************************************************************

lexA2_r.trimmed.seq      ATTGATGACGAAGTTACCGTTAAGCGCCTGAAAAAACAGGGCAATAAAGTCGAACTGTTG 513
lexA5_r.trimmed.seq      ATTGATGACGAAGTTACCGTTAAGCGCCTGAAAAAACAGGGCAATAAAGTCGAACTGTTG 532
lexA1_r.trimmed.seq      ATTGATGACGAAGTTACCGTTAAGCGCCTGAAAAAACAGGGCAATAAAGTCGAACTGTTG 513
lexA3_r.trimmed.seq      ATTGATGACGAAGTTACCGTTAAGCGCCTGAAAAAACAGGGCAATAAAGTCGAACTGTTG 524
EG10533                  ATTGATGACGAAGTTACCGTTAAGCGCCTGAAAAAACAGGGCAATAAAGTCGAACTGTTG 504
lexA4_r.trimmed.seq      ATTGATGACGAAGTTACCGTTAAGCGCCTGAAAAAACAGGGCAATAAAGTCGAACTGTTG 840
                         ************************************************************

lexA2_r.trimmed.seq      CCAGAAAATAGCGAGTTTAAACCAATTGTCGTTGACCTTCGTCAGCAGAGCTTCACCATT 573
lexA5_r.trimmed.seq      CCAGAAAATAGCGAGTTTAAACCAATTGTCGTTGACCTTCGTCAGCAGAGCTTCACCATT 592
lexA1_r.trimmed.seq      CCAGAAGATAGCGAGTTTAAACCAATTGTCGTTGACCTTCGTCAGCAGAGCTTCACCATT 573
lexA3_r.trimmed.seq      CCAGAAAATAGCGAGTTTAAACCAATTGTCGTTGACCTTCGTCAGCAGAGCTTCACCATT 584
EG10533                  CCAGAAAATAGCGAGTTTAAACCAATTGTCGTTGACCTTCGTCAGCAGAGCTTCACCATT 564
lexA4_r.trimmed.seq      CCAGAAAATAGCGAGTTTAAACCAATTGTCGTTGACCTTCGTCAGCAGAGCTTCACCATT 900
                         ****** ***************************************************

lexA2_r.trimmed.seq      GAAGGGCTGGCGGTTGGGGTTATTCGCAACGGCGACTGGCTGTAAAAGGGTTGGATCCTT 633
lexA5_r.trimmed.seq      GAAGGGCTGGCGGTTGGGGTTATTCGCAACGGCGACTGGCTGTAAAAGGGTTGGATCCTT 652
lexA1_r.trimmed.seq      GAAGGGCTGGCGGTTGGGGTTATTCGCAACGGCGACTGGCTGTAAAAGGGTTGGATCCTT 633
lexA3_r.trimmed.seq      GAAGGGCTGGCGGTTGGGGTTATTCGCAACGGCGACTGGCTGTAAAAGGGTTGGATCCTT 644
EG10533                  GAAGGGCTGGCGGTTGGGGTTATTCGCAACGGCGACTGGCTGTAA--------------- 609
lexA4_r.trimmed.seq      GAAGGGCTGGCGGTTGGGGTTATTCGCAACGGCGACTGGCTGTAAAAGGGCGAATTCGAA 960
                         *********************************************

lexA2_r.trimmed.seq      ATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACC 693
lexA5_r.trimmed.seq      ATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACC 712
lexA1_r.trimmed.seq      ATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACC 693
lexA3_r.trimmed.seq      ATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACC 704
EG10533                  ------------------------------------------------------------
lexA4_r.trimmed.seq      GCTGGCTTTGCG---------------------------------------------- 972


lexA2_r.trimmed.seq      ATGATGATGATGATGATGAGAACCCCCCATGGTTTATTCCTCCTTATTTAATCGATACAT 753
lexA5_r.trimmed.seq      ATGATGATGATGATGATGAGAACCCCCCATGGTTTATTCCTCCTTATTTAATCGATACAT 772
lexA1_r.trimmed.seq      ATGATGATGATGATGATGAGAACCCCCCATGGTTTATTCCTCCTTATTTAATCGATACAT 753
lexA3_r.trimmed.seq      ATGATGATGATGATGATGAGAACCCCCCATGGTTTATTCCTCCTTATTTAATCGATACAT 764
EG10533                  ------------------------------------------------------------
lexA4_r.trimmed.seq      ------------------------------------------------------------


lexA2_r.trimmed.seq      TAATATATACCTCTTTAATTTTTAATAATAAAGTTAATCGATAATTCCGGTCGAGTGCCC 813
lexA5_r.trimmed.seq      TAATATATACCTCTTTAATTTTTAATAATAAAGTTAATCGATAATTCCGGTCGAGTGCCC 832
lexA1_r.trimmed.seq      TAATATATACCTCTTTAATTTTTAATAATAAAGTTAATCGATAATTCCGGTCGAGTGCCC 813
lexA3_r.trimmed.seq      TAATATATACCTCTTTAATTTTTAATAATAAAGTTAATCGATAATTCCGGTCGAGTGCCC 824
EG10533                  ------------------------------------------------------------
lexA4_r.trimmed.seq      ------------------------------------------------------------


lexA2_r.trimmed.seq      ACACAGATTGTCTGATAAATTGTTAAAGAGCAGTGCCGCTTCGCTTTTTCTCAGCGGCGC 873
lexA5_r.trimmed.seq      ACACAGATTGTCTGATAAATTGTTAAAGAGCAGTGCCGCTTCGCTTTTTCTCAGCGGCGC 892
lexA1_r.trimmed.seq      ACACAGATTGTCTGATNAATTGTTAAAGAGCAGTGCCGCTTCGCTTTTTCTCAGCGGCGC 873
lexA3_r.trimmed.seq      ACACAGATTGTCTGATAAATTG---------------------------------------- 846
EG10533                  ------------------------------------------------------------
lexA4_r.trimmed.seq      ------------------------------------------------------------


lexA2_r.trimmed.seq      TGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACATTATACGAGCCGGA---- 929
lexA5_r.trimmed.seq      TGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACATTATACGAGCCGGATGAA 952
lexA1_r.trimmed.seq      TGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTC-ACACATTATAC------------ 920
lexA3_r.trimmed.seq      ------------------------------------------------------------
EG10533                  ------------------------------------------------------------
lexA4_r.trimmed.seq      ------------------------------------------------------------
```

```
lexA2_r.trimmed.seq      ---------------------
lexA5_r.trimmed.seq      TAATTGTCCACAGCTCATTTC 973
lexA1_r.trimmed.seq      ---------------------
lexA3_r.trimmed.seq      ---------------------
EG10533                  ---------------------
lexA4_r.trimmed.seq      ---------------------


CLUSTAL FORMAT for T-COFFEE Version_1.41, CPU=8.99 sec, SCORE=54, Nseq=3, Len=1526

nusA1 looks like it might be ok, but the nusA sequence is SO LONG that the read
couldn't cover it all

nusA2_r.trimmed.seq  CGCAACGCCAGCTTCGATTCGCCCTTATGAACAAAGAAATTTTGGCTGTAGTTGAAGCCG
nusA1_r.trimmed.seq  ------------------------------------------------------------
b3169                ---------------------------ATGAACAAAGAAATTTTGGCTGTAGTTGAAGCCG


nusA2_r.trimmed.seq  TATCCAATGAAAAGGCGCTACCTCGCGAGAAGATTTTCGAAGCATTGGAAAGCGCGCTGG
nusA1_r.trimmed.seq  ------------------------------------------------------------
b3169                TATCCAATGAAAAGGCGCTACCTCGCGAGAAGATTTTCGAAGCATTGGAAAGCGCGCTGG


nusA2_r.trimmed.seq  CGACAGCAACAAAGAAAAAATATGAACAAGAGATCGACGTCCGCGTACAGATCGATCGCA
nusA1_r.trimmed.seq  ------------------------------------------------------------
b3169                CGACAGCAACAAAGAAAAAATATGAACAAGAGATCGACGTCCGCGTACAGATCGATCGCA


nusA2_r.trimmed.seq  AAGGCGGTGATTTTGACACTTTCCGTCGCTGGTTAGTTGTTGATGAAGTCACCCAGCCGA
nusA1_r.trimmed.seq  ------------------------------------------------------------
b3169                AAAGCGGTGATTTTGACACTTTCCGTCGCTGGTTAGTTGTTGATGAAGTCACCCAGCCGA


nusA2_r.trimmed.seq  CCAAGGAAATCACCCTTGAAGCCGCACGTTATGAAGATGAAAGCCTGAACCTGGGCGATT
nusA1_r.trimmed.seq  ------------------------------------------------------------
b3169                CCAAGGAAATCACCCTTGAAGCCGCACGTTATGAAGATGAAAGCCTGAACCTGGGCGATT


nusA2_r.trimmed.seq  ACGTTGAAGATCAGATTGAGTCTGTTACCTTTGACCGTATCACTACCCAGACGGCAAAAC
nusA1_r.trimmed.seq  ------------------------------------------------------------
b3169                ACGTTGAAGATCAGATTGAGTCTGTTACCTTTGACCGTATCACTACCCAGACGGCAAAAC


nusA2_r.trimmed.seq  AGGTTATCGTGCAGAAAGTGCGTGAAGCCGAACGTGCGATGGTGGTTGATCAGTTCCGTG
nusA1_r.trimmed.seq  ------------------------------------------------------------
b3169                AGGTTATCGTGCAGAAAGTGCGTGAAGCCGAACGTGCGATGGTGGTTGATCAGTTCCGTG


nusA2_r.trimmed.seq  AACACGAAGGTGAAATCATCACCGGCGTGGTGAAAAAAGTAAACCGCGACAACATCTCTC
nusA1_r.trimmed.seq  ------------------------------------------------------------
b3169                AACACGAAGGTGAAATCATCACCGGCGTGGTGAAAAAAGTAAACCGCGACAACATCTCTC


nusA2_r.trimmed.seq  TGGATCTGGGCAACAACGCTGAAGCCGTGATCCTGCGCGAAGATATGCTGCCGCGTGAAA
nusA1_r.trimmed.seq  ------------------------------------------------------------
b3169                TGGATCTGGGCAACAACGCTGAAGCCGTGATCCTGCGCGAAGATATGCTGCCGCGTGAAA


nusA2_r.trimmed.seq  ACTTCCGCCCTGGCGACCGCGTTCGTGGCGTGCTCTATTCCGTTCGCCCGGAAGCGCGTG
nusA1_r.trimmed.seq  ---------------------------------------------CGCCCCGAAGCGCGTG
b3169                ACTTCCGCCCTGGCGACCGCGTTCGTGGCGTGCTCTATTCCGTTCGCCCGGAAGCGCGTG
                                                                  ***** **********

nusA2_r.trimmed.seq  GCGCGCAACTGTTCGTCACTCGTTCCAAGCCGGAAATGCTGATCGAACTGTTCCGTATTG
nusA1_r.trimmed.seq  GCGCGCAACTGTTCGTCACTCGTTCCAAGCCGGAAATGCTGATCGAACTGTTCCGTATTG
b3169                GCGCGCAACTGTTCGTCACTCGTTCCAAGCCGGAAATGCTGATCGAACTGTTCCGTATTG
                     ************************************************************

nusA2_r.trimmed.seq  AAGTGCCAGAAATCGGCGAAGAAGTGATTGAAATTAAAGCAGCGGCTCGCGATCCGGGTT
nusA1_r.trimmed.seq  AAGTGCCAGAAATCGGCGAAGAAGTGATTGAAATTAAAGCAGCGGCTCGCGATCCGGGTT
b3169                AAGTGCCAGAAATCGGCGAAGAAGTGATTGAAATTAAAGCAGCGGCTCGCGATCCGGGTT
                     ************************************************************

nusA2_r.trimmed.seq  CTCGTGCGAAAATCGCGGTGAAAACCAACGATAAACGTATCGATCCGGTAGGTGCTTGCG
nusA1_r.trimmed.seq  CTCGTGCGAAAATCGCGGTGAAAACCAACGATAAACGTATCGATCCGGTAGGTGCTTGCG
b3169                CTCGTGCGAAAATCGCGGTGAAAACCAACGATAAACGTATCGATCCGGTAGGTGCTTGCG
                     ************************************************************

nusA2_r.trimmed.seq  TANGTATGCGTGGCGCGCGTGTTCAGGCGGTGTCTACTGAACTGGGTGGCGAGCGTATCG
nusA1_r.trimmed.seq  TAGGTATGCGTGGCGCGCGTGTTCAGGCGGTGTCTACTGAACTGGGTGGCGAGCGTATCG
b3169                TAGGTATGCGTGGCGCGCGTGTTCAGGCGGTGTCTACTGAACTGGGTGGCGAGCGTATCG
                     ** ************************************************************

nusA2_r.trimmed.seq  ATATCGTCCTGTGGGATGATAACCCGGCGCAGTTCGTGATTAACGCAATGGCACCGGCAG
nusA1_r.trimmed.seq  ATATCGTCCTGTGGGATGATAACCCGGCGCAGTTCGTGATTAACGCAATGGCACCGGCAG
b3169                ATATCGTCCTGTGGGATGATAACCCGGCGCAGTTCGTGATTAACGCAATGGCACCGGCAG
                     ************************************************************

nusA2_r.trimmed.seq  ACGTTGCTTCTATCGTGGTGGATGAAGATAAACACACCATGGATATCGC-----------
nusA1_r.trimmed.seq  ACGTTGCTTCTATCGTGGTGGATGAAGATAAACACACCATGGATATCGCCGTTGAAGCCG
b3169                ACGTTGCTTCTATCGTGGTGGATGAAGATAAACACACCATGGATATCGCCGTTGAAGCCG
```

```
                  **************************************************

nusA2_r.trimmed.seq  ------------------------------------------------------------
nusA1_r.trimmed.seq  GTAACCTGGCGCAGGCGATTGGCCGTAACGGTCAGAACGTGCGTCTGGCTTCGCAGCTGA
b3169                GTAACCTGGCGCAGGCGATTGGCCGTAACGGTCAGAACGTGCGTCTGGCTTCGCAGCTGA


nusA2_r.trimmed.seq  ------------------------------------------------------------
nusA1_r.trimmed.seq  GCGGTTGGGAACTCAACGTGATGACCGTTGACGACCTGCAGGCTAAGCATCAGGCGGAAG
b3169                GCGGTTGGGAACTCAACGTGATGACCGTTGACGACCTGCAGGCTAAGCATCAGGCGGAAG


nusA2_r.trimmed.seq  ------------------------------------------------------------
nusA1_r.trimmed.seq  CGCACGCAGCGATCGACACCTTCACCAAATATCTCGACATCGACGAAGACTTCGCGACTG
b3169                CGCACGCAGCGATCGACACCTTCACCAAATATCTCGACATCGACGAAGACTTCGCGACTG


nusA2_r.trimmed.seq  ------------------------------------------------------------
nusA1_r.trimmed.seq  TTCTGGTAGAAGAAGGCTTCTCGACGCTGGAAGAATTGGCCTATGTGCCGATGAAAGAGC
b3169                TTCTGGTAGAAGAAGGCTTCTCGACGCTGGAAGAATTGGCCTATGTGCCGATGAAAGAGC


nusA2_r.trimmed.seq  ------------------------------------------------------------
nusA1_r.trimmed.seq  TGTTGGAAATCGAAGGCCTTGATGAGCCGACCGTTGAAGCACTGCGCGAGCGTGCTAAAA
b3169                TGTTGGAAATCGAAGGCCTTGATGAGCCGACCGTTGAAGCACTGCGCGAGCGTGCTAAAA


nusA2_r.trimmed.seq  ------------------------------------------------------------
nusA1_r.trimmed.seq  ATGCACTGGCCACCATTGCACAGGCCCAGGAAGAAAGCCTCGGTGATAACAAACCGGCTG
b3169                ATGCACTGGCCACCATTGCACAGGCCCAGGAAGAAAGCCTCGGTGATAACAAACCGGCTG


nusA2_r.trimmed.seq  ------------------------------------------------------------
nusA1_r.trimmed.seq  ACGATCTGCTGAACCTTGAAGGGGTAGATCGTGATTTGGCATTCAAACTGGCCGCCCGTG
b3169                ACGATCTGCTGAACCTTGAAGGGGTAGATCGTGATTTGGCATTCAAACTGGCCGCCCGTG


nusA2_r.trimmed.seq  ------------------------------------------------------------
nusA1_r.trimmed.seq  GCGTTTGTACGCTGGAAGATCTCGCCGAACAGGGCATTGATGATCTGGCTGATATCGAAG
b3169                GCGTTTGTACGCTGGAAGATCTCGCCGAACAGGGCATTGATGATCTGGCTGATATCGAAG


nusA2_r.trimmed.seq  ------------------------------------------------------------
nusA1_r.trimmed.seq  GGTTGACCGACGAAAAAGCCGGAGCACTGATTATGGCTGCCCGTAATATTTGCTGGTTCG
b3169                GGTTGACCGACGAAAAAGCCGGAGCACTGATTATGGCTGCCCGTAATATTTGCTGGTTCG


nusA2_r.trimmed.seq  ---------------------------
nusA1_r.trimmed.seq  GTGACGAAGCAAGGGCGAATTGAAGC
b3169                GTGACGAAGCGTAA------------


CLUSTAL W (1.82) multiple sequence alignment

pdhR7 is the only one not backwards.  Should sequence it again, as it might have mutation, but read is crappy.


EG11088                  ----------------------A-TGGCCTACAGCAAAATCCGCCAACCAAAACTCTCCGA 38
pdhR4_r.trimmed.seq      GCCAGCTTCGA-TTCGCCCTTA-TGGCCTACAGCAAAATCCGCCAACCAAAACTCTCCGA 58
pdhR1_r.trimmed.seq      -CCAGCTTCGAATTCGCCCTTAATGGCCTACAGCAAAATCCGCCAACCAAAACTCTCCGA 59
pdhR5_r.trimmed.seq      ------------TTCGCCCTTA-TGGCCTACAGCAAAATCCGCCA-CCAAAACTCTCCGA 46
pdhR2_r.trimmed.seq      ---AGCTTCGA-TTCGCCCTTA-TGGCCTACAGCAAAATCCGCCA-CCAAAACTCTCCGA 54
pdhR7_r.trimmed.seq      ------------------------------------------------------------


EG11088                  TGTGATTGAGCAGCAACTGGAGTTTTTGATCCTCGAAGGCACTCTCCGCCCGGGCGAAAA 98
pdhR4_r.trimmed.seq      TGTGATTGAGCAGCAACTGGAGTTTTTGATCCTCGAAGGCACTCTCCGCCCGGGCGAAAA 118
pdhR1_r.trimmed.seq      TGTGATTGAGCAGCAACTGGAGTTTTTGATCCTCGAAGGCACTCTCCGCCCGGGCGAAAA 119
pdhR5_r.trimmed.seq      TGTGATTGAGCAGCAACTGGAGTTTTTGATCCTCGAAGGCACTCTCCGCCCGGGCGAAAA 106
pdhR2_r.trimmed.seq      TGTGATTGAGCAGCAACTGGAGTTTTTGATCCTCGAAGGCACTCTCCGCCCGGGCGAAAA 114
pdhR7_r.trimmed.seq      ------------------------------------------------------------


EG11088                  ACTCCCACCGGAACGCGAACTGGCAAAACAGTTTGACGTCTCCCGTCCCTCCTTGCGTGA 158
pdhR4_r.trimmed.seq      ACTCCCACCGGAACGCGAACTGGCAAAACAGTTTGACGTCTCCCGTCCCTCCTTGCGTGA 178
pdhR1_r.trimmed.seq      ACTCCCACCGGAACGCGAACTGGCAAAACAGTTTGACGTCTCCCGTCCCTCCTTGCGTGA 179
pdhR5_r.trimmed.seq      ACTCCCACCGGAACGCGAACTGGCAAAACAGTTTGACGTCTCCCGTCCCTCCTTGCGTGA 166
pdhR2_r.trimmed.seq      ACTCCCACCGGAACGCGAACTGGCAAAACAGTTTGACGTCTCCCGTCCCTCCTTGCGTGA 174
pdhR7_r.trimmed.seq      ----CCACCGGAACGCGAACTGGCAAAACAGTTGGACGTCTCCCGTCCCTCCTTGCGTGA 56
                             *************************** ************************


EG11088                  GGCGATTCAACGTCTCGAAGCGAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTT 218
pdhR4_r.trimmed.seq      GGCGATTCAACGTCTCGAAGCGAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTT 238
pdhR1_r.trimmed.seq      GGCGATTCAACGTCTCGAAGCGAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTT 239
pdhR5_r.trimmed.seq      GGCGATTCAACGTCTCGAAGCGAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTT 226
pdhR2_r.trimmed.seq      GGCGATTCAACGTCTCGAAGCGAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTT 234
pdhR7_r.trimmed.seq      GGCGATTCAACGTCTCGAAGCGAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGGCACTTT 116
                         ************************************************************


EG11088                  TGTCCAGAGCAGCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGCTCTCCGACCA 278
pdhR4_r.trimmed.seq      TGTCCAGAGCAGCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGCTCTCCGACCA 298
pdhR1_r.trimmed.seq      TGTCCAGAGCAGCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGCTCTCCGACCA 299
```

```
pdhR5_r.trimmed.seq      TGTCCAGAGCAGCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGCTCTCCGACCA 286
pdhR2_r.trimmed.seq      TGTCCAGAGCAGCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGCTCTCCGACCA 294
pdhR7_r.trimmed.seq      TGTCCAGAGCAGCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGCTCTCCGACCA 176
                         ************************************************************

EG11088                  TCCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCCCTGGAAGGTATCGCCGCTTA 338
pdhR4_r.trimmed.seq      TCCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCCCTGGAAGGTATCGCCGCTTA 358
pdhR1_r.trimmed.seq      TCCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCCCTGGAAGGTATCGCCGCTTA 359
pdhR5_r.trimmed.seq      TCCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCCCTGGAAGGTATCGCCGCTTA 346
pdhR2_r.trimmed.seq      TCCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCCCTGGAAGGTATCGCCGCTTA 354
pdhR7_r.trimmed.seq      TCCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCCCTGGAAGGTATCGCCGCTTA 236
                         ************************************************************

EG11088                  TTACGCCGCGCTGCGTAGTACCGATGAAGACAAGGAACGCATCCGTGAACTCCACCACGC 398
pdhR4_r.trimmed.seq      TTACGCCGCGCTGCGTAGTACCGATGAAGACAAGGAACGCATCCGTGAACTCCACCACGC 418
pdhR1_r.trimmed.seq      TTACGCCGCGCTGCGTAGTACCGATGAAGACAAGGAACGCATCCGTGAACTCCACCACGC 419
pdhR5_r.trimmed.seq      TTACGCCGCGCTGCGTAGTACCGATGAAGACAAGGAACGCATCCGTGAACTCCACCACGC 406
pdhR2_r.trimmed.seq      TTACGCCGCGCTGCGTAGTACCGATGAAGACAAGGAACGCATCCGTGAACTCCACCACGC 414
pdhR7_r.trimmed.seq      TTACGCCGCGCTGCGTAGTACCGATGAAGACAAGGAACGCATCCGTGAACTCCACCACGC 296
                         ************************************************************

EG11088                  CATAGAGCTGGCGCAGCAGTCTGGCGATCTGGACGCGGAATCAAACGCCGTACTCCAGTA 458
pdhR4_r.trimmed.seq      CATAGAGCTGGCGCAGCAGTCTGGCGATCTGGACGCGGAATCAAACGCCGTACTCCAGTA 478
pdhR1_r.trimmed.seq      CATAGAGCTGGCGCAGCAGTCTGGCGATCTGGACGCGGAATCAAACGCCGTACTCCAGTA 479
pdhR5_r.trimmed.seq      CATAGAGCTGGCGCAGCAGTCTGGCGATCTGGACGCGGAATCAAACGCCGTACTCCAGTA 466
pdhR2_r.trimmed.seq      CATAGAGCTGGCGCAGCAGTCTGGCGATCTGGACGCGGAATCAAACGCCGTACTCCAGTA 474
pdhR7_r.trimmed.seq      CATAGAGCTGGCGCAGCAGTCTGGCGATCTGGACGCGGAATCAAACGCCGTACTCCAGTA 356
                         ************************************************************

EG11088                  TCAGATTGCCGTCACCGAAGCGGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTAT 518
pdhR4_r.trimmed.seq      TCAGATTGCCGTCACCGAAGCGGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTAT 538
pdhR1_r.trimmed.seq      TCAGATTGCCGTCACCGAAGCGGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTAT 539
pdhR5_r.trimmed.seq      TCAGATTGCCGTCACCGAAGCGGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTAT 526
pdhR2_r.trimmed.seq      TCAGATTGCCGTCACCGAAGCGGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTAT 534
pdhR7_r.trimmed.seq      TCAGATTGCCGTCACCGAAGCGGCCCACAATGTGGTTCTGCTTCATCTGCTAAGGTGTAT 416
                         ************************************************************

EG11088                  GGAGCCGATGTTGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATTCGCGTCGCGA 578
pdhR4_r.trimmed.seq      GGAGCCGATGTTGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATTCGCGTCGCGA 598
pdhR1_r.trimmed.seq      GGAGCCGATGTTGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATTCGCGTCGCGA 599
pdhR5_r.trimmed.seq      GGAGCCGATGTTGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATTCGCGTCGCGA 586
pdhR2_r.trimmed.seq      GGAGCCGATGTTGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATTCGCGTCGCGA 594
pdhR7_r.trimmed.seq      GGAGCCGATGTTGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATTCGCGTCGCGA 476
                         ************************************************************

EG11088                  GATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTTGAAGCGATTATGGCCGGTAA 638
pdhR4_r.trimmed.seq      GATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTTGAAGCGATTATGGCCGGTAA 658
pdhR1_r.trimmed.seq      GATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTTGAAGCGATTATGGCCGGTAA 659
pdhR5_r.trimmed.seq      GATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTTGAAGCGATTATGGCCGGTAA 646
pdhR2_r.trimmed.seq      GATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTTGAAGCGATTATGGCCGGTAA 654
pdhR7_r.trimmed.seq      GATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTTGAAGCGATTATGGCCGGTAA 536
                         ************************************************************

EG11088                  GCCGGAAGAAGCGCGCCGAAGCATCGCATCGCCATCTGGCCTTTATCGAAGAAATTTTGCT 698
pdhR4_r.trimmed.seq      GCCGGAAGAAGCGCGCGAAGCATCGCATCGCCATCTGGCCTTTATCGAAGAAATTTTGCT 718
pdhR1_r.trimmed.seq      GCCGGAAGAAGCGCGCGAAGCATCGCATCGCCATCTGGCCTTTATCGAAGAAATTTTGCT 719
pdhR5_r.trimmed.seq      GCCGGAAGAAGCGCGCCGAAGCATCGCATCGCCATCTGGCCTTTATCGAAGAAATTTTGCT 706
pdhR2_r.trimmed.seq      GCCGGAAGAAGCGCGCGAAGCATCGCATCGCCATCTGGCCTTTATCGAAGAAATTTTGCT 714
pdhR7_r.trimmed.seq      GCCGGAAGAAGCGCGCGAAGCATCGCATCGCCATCTGGCCTTTATCGAAGAAATTTTGCT 596
                         ************************************************************

EG11088                  CGACAGAAGTCGTGAAGAGAGCCGCCGTGAGCGTTCTCTGCGTCGTCTGGAGCAACGAAA 758
pdhR4_r.trimmed.seq      CGACAGAAGTCGTGAAGAGAGCCGCCGTGAGCGTTCTCTGCGTCGTCTGGAGCAACGAAA 778
pdhR1_r.trimmed.seq      CGACAGAAGTCGTGAAGAGAGCCGCCGTGAGCGTTCTCTGCGTCGTCTGGAGCAACGAAA 779
pdhR5_r.trimmed.seq      CGACAGAAGTCGTGAAGAGAGCCGCCGTGAGCGTTCTCTGCGTCGTCTGGAGCAACGAAA 766
pdhR2_r.trimmed.seq      CGACAGAAGTCGTGAAGAGAGCCGCCGTGAGCGTTCTCTGCGTCGTCTGGAGCAACGAAA 774
pdhR7_r.trimmed.seq      CGACAGAAGTCGTGAAGAGAGCCGCCGTGAGCGTTCTCTGCGTCGTCTGGAGCAACGAAA 656
                         ************************************************************

EG11088                  GAATTAG------------------------------------------------------ 765
pdhR4_r.trimmed.seq      GAATTAGAAGGGTTGGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTC 838
pdhR1_r.trimmed.seq      GAATTAGAAGGGTTGGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTC 839
pdhR5_r.trimmed.seq      GAATTAGAAGGGTTGGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTC 826
pdhR2_r.trimmed.seq      GAATTAGAAGGGTTGGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTC 834
pdhR7_r.trimmed.seq      GAAT-AGAAGGGC-GAATCGAAGCTGGCTGTTGCGAA----------------------- 691
                         **** **

EG11088                  ------------------------------------------------------------
pdhR4_r.trimmed.seq      CACCAGTCATGCTAGCCATACCATGATGATGATGATGATGAGAACCCCCCATGGTTTATT 898
pdhR1_r.trimmed.seq      CACCAGTCATGCTAGCCATACCATGATGATGATGATGATGAGAACCCCCCATGGTTTATT 899
pdhR5_r.trimmed.seq      CACCAGTCATGCTAGCCATACCATGATGATGATGATGATGAGAACCCCCCATGGTTTATT 886
pdhR2_r.trimmed.seq      CACCAGTCATGCTAGCCATACCATGATGATGATGATGATGAGAACCCCCCATGGTTTATT 894
pdhR7_r.trimmed.seq      ------------------------------------------------------------

EG11088                  ------------------------------------------------------------
pdhR4_r.trimmed.seq      CCTCCTTATTTTATCGATACATTAATATATACCTCTTTAATTTT---------------- 942
pdhR1_r.trimmed.seq      CCTCCTTATTT-ATCGATACATAAATATATA----------------------------- 929
pdhR5_r.trimmed.seq      CC--------------------------------------------------------- 888
pdhR2_r.trimmed.seq      CCTCCTTTATTTATCGATACATTAATATATACCTCTTTATTTTTAATAATAAGTTAATCG 954
pdhR7_r.trimmed.seq      ------------------------------------------------------------
```

```
CLUSTAL W (1.82) multiple sequence alignment

YES!!! (the deletion is plainly a miscall if you see the chromatagraph)


rhaR1_f.trimmed.seq      ACCCTTATGGCTTTCTGCA-TAACGCGAATCTTCTCAACGTATTTGTACGCCATATTGCG 59
EG10842                  ------ATGGCTTTCTGCAATAACGCGAATCTTCTCAACGTATTTGTACGCCATATTGCG 54
rhaR1_r.trimmed.seq      ----------------GCAATAACGCGAATCTTCTCAACGTATTTGTACGCCATATTGCG 44
                                        *** ************************************

rhaR1_f.trimmed.seq      AATAATCAACTTCGTTCTCTGGCCGAGGTAGCCACGGTGGCGCATCAGTTAAAACTTCTC 119
EG10842                  AATAATCAACTTCGTTCTCTGGCCGAGGTAGCCACGGTGGCGCATCAGTTAAAACTTCTC 114
rhaR1_r.trimmed.seq      AATAATCAACTTCGTTCTCTGGCCGAGGTAGCCACGGTGGCGCATCAGTTAAAACTTCTC 104
                         ************************************************************

rhaR1_f.trimmed.seq      AAAGATGATTTTTTTGCCAGCGACCAGCAGGCAGTCGCTGTGGCTGACCGTTATCCGCAA 179
EG10842                  AAAGATGATTTTTTTGCCAGCGACCAGCAGGCAGTCGCTGTGGCTGACCGTTATCCGCAA 174
rhaR1_r.trimmed.seq      AAAGATGATTTTTTTGCCAGCGACCAGCAGGCAGTCGCTGTGGCTGACCGTTATCCGCAA 164
                         ************************************************************

rhaR1_f.trimmed.seq      GATGTCTTTGCTGAACATACACATGATTTTTGTGAGCTGGTGATTGTCTGGCGCGGTAAT 239
EG10842                  GATGTCTTTGCTGAACATACACATGATTTTTGTGAGCTGGTGATTGTCTGGCGCGGTAAT 234
rhaR1_r.trimmed.seq      GATGTCTTTGCTGAACATACACATGATTTTTGTGAGCTGGTGATTGTCTGGCGCGGTAAT 224
                         ************************************************************

rhaR1_f.trimmed.seq      GGCCTGCATGTACTCAACGATCGCCCTTATCGCATTACCCGTGGCGATCTCTTTTACATT 299
EG10842                  GGCCTGCATGTACTCAACGATCGCCCTTATCGCATTACCCGTGGCGATCTCTTTTACATT 294
rhaR1_r.trimmed.seq      GGCCTGCATGTACTCAACGATCGCCCTTATCGCATTACCCGTGGCGATCTCTTTTACATT 284
                         ************************************************************

rhaR1_f.trimmed.seq      CATGCTGACGATAAACACTCCTACGCTTCCGTTAACGATCTGGTTTTGCAGAATATTATT 359
EG10842                  CATGCTGACGATAAACACTCCTACGCTTCCGTTAACGATCTGGTTTTGCAGAATATTATT 354
rhaR1_r.trimmed.seq      CATGCTGACGATAAACACTCCTACGCTTCCGTTAACGATCTGGTTTTGCAGAATATTATT 344
                         ************************************************************

rhaR1_f.trimmed.seq      TATTGCCCGGAGCGTCTGAAGCTGAATCTTGACTGGCAGGGGGCGATTCCGGGATTTAAC 419
EG10842                  TATTGCCCGGAGCGTCTGAAGCTGAATCTTGACTGGCAGGGGGCGATTCCGGGATTTAAC 414
rhaR1_r.trimmed.seq      TATTGCCCGGAGCGTCTGAAGCTGAATCTTGACTGGCAGGGGGCGATTCCGGGATTTAAC 404
                         ************************************************************

rhaR1_f.trimmed.seq      GCCAGCGCAGGGCAACCACACTGGCGCTTAGGTAGCATGGGGATGGCGCAGGCGCGGCAG 479
EG10842                  GCCAGCGCAGGGCAACCACACTGGCGCTTAGGTAGCATGGGGATGGCGCAGGCGCGGCAG 474
rhaR1_r.trimmed.seq      GCCAGCGCAGGGCAACCACACTGGCGCTTAGGTAGCATGGGGATGGCGCAGGCGCGGCAG 464
                         ************************************************************

rhaR1_f.trimmed.seq      GTTATCGGTCAGCTTGAGCATGAAAGTAGTCAGCATGTGCCGTTTGCTAACGAAATGGCT 539
EG10842                  GTTATCGGTCAGCTTGAGCATGAAAGTAGTCAGCATGTGCCGTTTGCTAACGAAATGGCT 534
rhaR1_r.trimmed.seq      GTTATCGGTCAGCTTGAGCATGAAAGTAGTCAGCATGTGCCGTTTGCTAACGAAATGGCT 524
                         ************************************************************

rhaR1_f.trimmed.seq      GAGTTGCTGTTCGGGCAGTTGGTGATGTTGCTGAATCGCCATCGTTACACCAGTGATTCG 599
EG10842                  GAGTTGCTGTTCGGGCAGTTGGTGATGTTGCTGAATCGCCATCGTTACACCAGTGATTCG 594
rhaR1_r.trimmed.seq      GAGTTGCTGTTCGGGCAGTTGGTGATGTTGCTGAATCGCCATCGTTACACCAGTGATTCG 584
                         ************************************************************

rhaR1_f.trimmed.seq      TTGCCGCCAACATCCAGCGAAACGTTGCTGGATAAGCTGATTACCCGGCTGGCGGCTAGC 659
EG10842                  TTGCCGCCAACATCCAGCGAAACGTTGCTGGATAAGCTGATTACCCGGCTGGCGGCTAGC 654
rhaR1_r.trimmed.seq      TTGCCGCCAACATCCAGCGAAACGTTGCTGGATAAGCTGATTACCCGGCTGGCGGCTAGC 644
                         ************************************************************

rhaR1_f.trimmed.seq      CTGAAAAGTCCCTTTGCGCTGGATAAATTTTGTGATGAGGCATCGTGCAGTGAGCGCGTT 719
EG10842                  CTGAAAAGTCCCTTTGCGCTGGATAAATTTTGTGATGAGGCATCGTGCAGTGAGCGCGTT 714
rhaR1_r.trimmed.seq      CTGAAAAGTCCCTTTGCGCTGGATAAATTTTGTGATGAGGCATCGTGCAGTGAGCGCGTT 704
                         ************************************************************

rhaR1_f.trimmed.seq      TTGCGTCAGCAATTTCGCCAGCAGACTGGAATGACCATCAATCAATATCTGCGACAGGTC 779
EG10842                  TTGCGTCAGCAATTTCGCCAGCAGACTGGAATGACCATCAATCAATATCTGCGACAGGTC 774
rhaR1_r.trimmed.seq      TTGCGTCAGCAATTTCGCCAGCAGACTGGAATGACCATCAATCAATATCTGCGACAGGTC 764
                         ************************************************************

rhaR1_f.trimmed.seq      AGAGTGTGTCATGCGCAATATCTTCTCCAGCATAGCCGCCTGTTAATCAGTGATATTTCG 839
EG10842                  AGAGTGTGTCATGCGCAATATCTTCTCCAGCATAGCCGCCTGTTAATCAGTGATATTTCG 834
rhaR1_r.trimmed.seq      AGAGTGTGTCATGCGCAATATCTTCTCCAGCATAGCCGCCTGTTAATCAGTGATATTTCG 824
                         ************************************************************

rhaR1_f.trimmed.seq      ACCGAATGTGGCTTTGAAGATAGTAACTATTTTTCGGTGGTGTTTACCCGGGAAACCGGG 899
EG10842                  ACCGAATGTGGCTTTGAAGATAGTAACTATTTTTCGGTGGTGTTTACCCGGGAAACCGGG 894
rhaR1_r.trimmed.seq      ACCGAATGTGGCTTTGAAGATAGTAACTATTTTTCGGTGGTGTTTACCCGGGAAACCGGG 884
                         ************************************************************

rhaR1_f.trimmed.seq      ATGACGCCCAGCCAGTGGCGTCATC----------------------------------- 924
EG10842                  ATGACGCCCAGCCAGTGGCGTCATCTCAATTCGCAGAAAGATTAA--------------- 939
rhaR1_r.trimmed.seq      ATGACGCCCAGCCAGTGGCGTCATCTCAATTCGCAGAAAGATAAAAGGGCGAATCGAAGC 944
                         *************************

rhaR1_f.trimmed.seq      ----------
EG10842                  ----------
rhaR1_r.trimmed.seq      TGGCGTTGCG 954


CLUSTAL W (1.82) multiple sequence alignment
```

```
NONE backwards, but most have funny things like inserts or mutations.
ydaK4 looks best, but it has one N.  If you check chromatograph it looks strange, but like
the others that were called correctly.


ydaK2_r.trimmed.seq      ------------------------------------------------------------
ydaK4_r.trimmed.seq      CATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGAT 60
G6671                    ------------------------------------------------------------
ydaK3_r.trimmed.seq      --------TCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGAT 52


ydaK2_r.trimmed.seq      ----------------------------------------------------ATTCATCAA 9
ydaK4_r.trimmed.seq      CTGTACGACGATGACGATAAGATTCCACCCTTATGGGCTTTTCAGGTAAAAATTCATCAA 120
G6671                    ---------------------------------ATGGCTTTTCAGGTAAAAATTCATCAA 27
ydaK3_r.trimmed.seq      CTGTACGACGATGACGATAAGGATTCCACCCTTATGGCTTTTCAGGTAAAAATTCATCAA 112
                                                                         *********

ydaK2_r.trimmed.seq      ATTCGGGCTTTGTTTGAAGTGGCTCGTCAGGGCAGCATTCGCGGAGCGAGCCGAATGTTG 69
ydaK4_r.trimmed.seq      ATTCGGGCTTTTGTTGAAGTGGCTCGTCAGGGCAGCATTCGCGGAGCGAGCCGAATGTTG 180
G6671                    ATTCGGGCTTTTGTTGAAGTGGCTCGTCAGGGCAGCATTCGCGGAGCGAGCCGAATGTTG 87
ydaK3_r.trimmed.seq      ATTCGGGCTTTTGTTGAAGTGGCTCGTCAGGGCAGCATTCGCGGAGCGAGCCGAATGTTG 172
                         **********  **************************************************

ydaK2_r.trimmed.seq      AATATGTCGCAACCGGCACTGAGTAAATCTATTCAGGAGCTAGAAGAAGGGTTAGCGGCG 129
ydaK4_r.trimmed.seq      AATATGTCGCAACCGGCACTGAGTAAATCTATTCAGGAGCTAGAAGAAGGGTTAGCGGCG 240
G6671                    AATATGTCGCAACCGGCACTGAGTAAATCTATTCAGGAGCTAGAAGAAGGGTTAGCGGCG 147
ydaK3_r.trimmed.seq      AATATGTCGCAACCGGCACTGAGTAAATCTATTCAGGAGCTAGAAGAAGGGTTAGCGGCG 232
                         ************************************************************

ydaK2_r.trimmed.seq      CAACTCTTTTTTCGCCGTAGTAAAGGCGTGACGTTAACTGATGCCGGTGAAAGTTTTTAT 189
ydaK4_r.trimmed.seq      CAACTCTTTNTTCGCCGTAGTAAAGGCGTGACGTTAACTGATGCCGGTGAAAGTTTTTAT 300
G6671                    CAACTCTTTTTTCGCCGTAGTAAAGGCGTGACGTTAACTGATGCCGGTGAAAGTTTTTAT 207
ydaK3_r.trimmed.seq      CAACTCTTTTTTCGCCGTAGTAAAGGCGTGACGTTAACTGATGCCGGTGAAAGTTTTTAT 292
                         ********* ***************************************************

ydaK2_r.trimmed.seq      CAGCACGCCAGTCTAATTCTTGAAGAGCTGCGCGCAGCCCAAGAGGATATTCGCCAACGA 249
ydaK4_r.trimmed.seq      CAGCACGCCAGTCTAATTCTTGAAGAGCTGCGCGCAGCCCAAGAGGATATTCGCCAACGA 360
G6671                    CAGCACGCCAGTCTAATTCTTGAAGAGCTGCGCGCAGCCCAAGAGGATATTCGCCAACGA 267
ydaK3_r.trimmed.seq      CAGCACGCCAGTCTAATTCTTGAAGAGCTGCGCGCAGCCCAAGAGGATATTCGCCAACGA 352
                         ************************************************************

ydaK2_r.trimmed.seq      CAAGGGCAACTGGCAGGGCAGATTAATATCGGCATGGGGGCCAGTATTTCCCGCAGTCTG 309
ydaK4_r.trimmed.seq      CAAGGGCAACTGGCAGGGCAGATTAATATCGGCATGGGGGCCAGTATTTCCCGCAGTCTG 420
G6671                    CAAGGGCAACTGGCAGGGCAGATTAATATCGGCATGGGGGCCAGTATTTCCCGCAGTCTG 327
ydaK3_r.trimmed.seq      CAAGGGCAACTGGCAGGGCAGATTAATATCGGCATGGGGGCCAGTATTTCCCGCAGTCTG 412
                         ************************************************************

ydaK2_r.trimmed.seq      ATGCCAGCTGTCATATCTCGTTTTCATCAGCAGCATCCGCAGGTAAAAGTACGCATTATG 369
ydaK4_r.trimmed.seq      ATGCCAGCTGTCATATCTCGTTTTCATCAGCAGCATCCGCAGGTAAAAGTACGCATTATG 480
G6671                    ATGCCAGCTGTCATATCTCGTTTTCATCAGCAGCATCCGCAGGTAAAAGTACGCATTATG 387
ydaK3_r.trimmed.seq      ATGCCAGCTGTCATATCTCGTTTTCATCAGCAGTATCCGCAGGTAAAAGTACGCATTATG 472
                         ********************************** *************************

ydaK2_r.trimmed.seq      GAAGGGCAACTGGTGTCGATGATTAATGAATTGCGTCAGGGAGAATTGGATTTCACCATC 429
ydaK4_r.trimmed.seq      GAAGGGCAACTGGTGTCGATGATTAATGAATTGCGTCAGGGAGAATTGGATTTCACCATC 540
G6671                    GAAGGGCAACTGGTGTCGATGATTAATGAATTGCGTCAGGGAGAATTGGATTTCACCATC 447
ydaK3_r.trimmed.seq      GAAGGGCAACTGGTGTCGATGATTAATGAATTGCGTCAGGGAGAATTGGATTTCACCATC 532
                         ************************************************************

ydaK2_r.trimmed.seq      AATACCTATTATCAGGGACCGTACGACCACGAATTTACTTTTGAGAAATTACTGGAAAAG 489
ydaK4_r.trimmed.seq      AATACCTATTATCAGGGACCGTACGACCACGAATTTACTTTTGAGAAATTACTGGAAAAG 600
G6671                    AATACCTATTATCAGGGACCGTACGACCACGAATTTACTTTTGAGAAATTACTGGAAAAG 507
ydaK3_r.trimmed.seq      AATACCTATTATCAGGGACCGTACGACCACGAATTTACTTTTGAGAAATTACTGGAAAAG 592
                         ************************************************************

ydaK2_r.trimmed.seq      CAATTCGCGATCTTTTGCCGCCCCGGGACACCCCGCCATTGGTGCCCGTTCGATCAAACAG 549
ydaK4_r.trimmed.seq      CAATTCGCGATCTTTTGCCGCCCGGGACACCCCGCCATTGGTGCCCGTTCGATCAAACAG 660
G6671                    CAATTCGCGATCTTTTGCCGCCCGGGACACCCCGCCATTGGTGCCCGTTCGATCAAACAG 567
ydaK3_r.trimmed.seq      CAATTCGCGATCTTTTGCCGCCCGGGACACCCCGCCATTGGTGCCCGTTCGATCAAACAG 652
                         ************************************************************

ydaK2_r.trimmed.seq      TTACTGGATTACAGCTGGACAATGCCGACGCCACACGGCAGCTACTACAAACAGTTGAGT 609
ydaK4_r.trimmed.seq      TTACTGGATTACAGCTGGACAATGCCGACGCCACACGGCAGCTACTACAAACAGTTGAGT 720
G6671                    TTACTGGATTACAGCTGGACAATGCCGACGCCACACGGCAGCTACTACAAACAGTTGAGT 627
ydaK3_r.trimmed.seq      TTACTGGATTACAGCTGGACAATGCCGACGCCACACGGCAGCTACTACAAACAGTTGAGT 712
                         ************************************************************

ydaK2_r.trimmed.seq      GAATTGCTTGACGATCAGGCGCAAACGCCACAGGTCGGTGTAGTCTGCGAGACGTTCTCA 669
ydaK4_r.trimmed.seq      GAATTGCTTGACGATCAGGCGCAAACGCCACAGGTCGGTGTAGTCTGCGAGACGTTCTCA 780
G6671                    GAATTGCTTGACGATCAGGCGCAAACGCCACAGGTCGGTGTAGTCTGCGAGACGTTCTCA 687
ydaK3_r.trimmed.seq      GAATTGCTTGACGATCAGGCGCAAACGCCACAGGTCGGTGTAGTCTGCGAGACGTTCTCA 772
                         ************************************************************

ydaK2_r.trimmed.seq      GCCTGTATCAGTCTGGTGGCAAAAAGCGATTTTCTCAGCAAACTGCCTGAAGAAATGGGC 729
ydaK4_r.trimmed.seq      GCCTGTATCAGTCTGGTGGCAAAAAGCGATTTTCTCAGCAAACTGCCTGAAGAAATGGGC 840
G6671                    GCCTGTATCAGTCTGGTGGCAAAAAGCGATTTTCTCAGCAAACTGCCTGAAGAAATGGGC 747
ydaK3_r.trimmed.seq      GCCTGTATCAGTCTGGTGGCAAAAAGCGATTTTCTCAGCAAACTGCCTGAAGAAATGGGC 832
                         ************************************************************

ydaK2_r.trimmed.seq      TGCGATCCCTTGCACGGACAGGGGCTGGTGATGTTGCCGGTTAGCGAAATTTTACCGAAA 789
ydaK4_r.trimmed.seq      TGCGATCCCTTGCACGGACAGGGGCTGGTGATGTTGCCGGTTAGCGAAATTTTACCGAAA 900
G6671                    TGCGATCCCTTGCACGGACAGGGGCTGGTGATGTTGCCGGTTAGCGAAATTTTACCGAAA 807
```

```
ydaK3_r.trimmed.seq      TGCGATCCCTTGCACGGACAGGGGCTGGTGATGTTGCCGGTTAGCGAAATTTTACCGAAA 892
                         ************************************************************

ydaK2_r.trimmed.seq      GCGGCCTATTATTTGATTCAGCGGCGTGATAGTCGCCAGACACCACTGACCGCGTCATTA 849
ydaK4_r.trimmed.seq      GCGGCCTATTATTTGATTCAGCGGCGTGATAGTCGCCAGACACCACTGACCGCGTCATTA 960
G6671                    GCGGCCTATTATTTGATTCAGCGGCGTGATAGTCGCCAGACACCACTGACCGCGTCATTA 867
ydaK3_r.trimmed.seq      GCGGCCTATTATTTGATTCAGCGGCGTGATAGTCGCCAGACACCACTGACCGCGTCATTA 952
                         ************************************************************

ydaK2_r.trimmed.seq      ATCACGCAATTCCGGCGAGAATGCGGCTATCTGCAAAGTTAAAGGGCGAATTCGAAGC-- 907
ydaK4_r.trimmed.seq      ATCACGCAATTCCGGCGAGAATGCGGCTATCTGCAAAGTTAAAGGGCGAATTCGAAGC-- 1018
G6671                    ATCACGCAATTCCGGCGAGAATGCGGCTATCTGCAAAGTTAA------------------ 909
ydaK3_r.trimmed.seq      ATCACGCAATTCCGGCGAGAATGCGGCTATCTGCAAAGTAAAAGGGCGAATCGAAGCTGG 1012
                         ************************************** **

ydaK2_r.trimmed.seq      ------
ydaK4_r.trimmed.seq      ------
G6671                    ------
ydaK3_r.trimmed.seq      CGTTGC 1018
```

Brief Conclusions: Finally some good news on the cloning front!!!!! definitely correctly cloned genes: fecI, fliA, lexA, gfp, ydaK. genes the are most likely correct but need another resequencing effort to make sure: nusA and pdhR.

```
CLUSTAL W (1.82) multiple sequence alignment


EG11088                  ------------------------------------------------------------
pdhR_7_R.trimmed.seq     CAGACAATCTGTGTGGGCACTCGACCGGAATTATCGATTACTTTATTATTAAAAATTAAA 60


EG11088                  ------------------------------------------------------------
pdhR_7_R.trimmed.seq     GAGGTATATATTAATGTATCGATTAAATAAGGAGGAATAAACCATGGGGGGTTCTCATCA 120


EG11088                  ------------------------------------------------------------
pdhR_7_R.trimmed.seq     TCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTA 180


EG11088                  ----------------------------ATGGCCTACAGCAAAATCCGCCAACCAAAACT 32
pdhR_7_R.trimmed.seq     CGACGATGACGATAAGGATCCAACCCTTATGGCCTACAGCAAAATCCGCCAACCAAAACT 240
                                                     ******************************


EG11088                  CTCCGATGTGATTGAGCAGCAACTGGAGTTTTTGATCCTCGAAGGCACTCTCCGCCCGGG 92
pdhR_7_R.trimmed.seq     CTCCGATGTGATTGAGCAGCAACTGGAGTTTTTGATCCTCGAAGGCACTCTCCGCCCGGG 300
                         ************************************************************

EG11088                  CGAAAAACTCCCACCGGAACGCGAACTGGCAAAACAGTTTGACGTCTCCCGTCCCTCCTT 152
pdhR_7_R.trimmed.seq     CGAAAAACTCCCACCGGAACGCGAACTGGCAAAACAGTTTGACGTCTCCCGTCCCTCCTT 360
                         ************************************************************

EG11088                  GCGTGAGGCGATTCAACGTCTCGAAGCGAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGG 212
pdhR_7_R.trimmed.seq     GCGTGAGGCGATTCAACGTCTCGAAGCGAAGGGCTTGTTGCTTCGTCGCCAGGGTGGCGG 420
                         ************************************************************

EG11088                  CACTTTTGTCCAGAGCAGCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGCTCTC 272
pdhR_7_R.trimmed.seq     CACTTTTGTCCAGAGCAGCCTATGGCAAAGCTTCAGCGATCCGCTGGTGGAGCTGCTCTC 480
                         ************************************************************

EG11088                  CGACCATCCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCCCTGGAAGGTATCGC 332
pdhR_7_R.trimmed.seq     CGACCATCCTGAGTCACAGTATGACTTGCTCGAAACACGACACGCCCTGGAAGGTATCGC 540
                         ************************************************************

EG11088                  CGCTTATTACGCCGCGCTGCGTAGTACCGATGAAGACAAGGAACGCATCCGTGAACTCCA 392
pdhR_7_R.trimmed.seq     CGCTTATTACGCCGCGCTGCGTAGTACCGATGAAGACAAGGAACGCATCCGTGAACTCCA 600
                         ************************************************************

EG11088                  CCACGCCATAGAGCTGGCGCAGCAGTCTGGCGATCTGGACGCGGAATCAAACGCCGTACT 452
pdhR_7_R.trimmed.seq     CCACGCCATAGAGCTGGCGCAGCAGTCTGGCGATCTGGACGCGGAATCAAACGCCGTACT 660
                         ************************************************************

EG11088                  CCAGTATCAGATTGCCGTCACCGAAGCGGCCCACAATGTGGTTCTGCTTCATCTGCTAAG 512
pdhR_7_R.trimmed.seq     CCAGTATCAGATTGCCGTCACCGAAGCGGCCCACAATGTGGTTCTGCTTCATCTGCTAAG 720
                         ************************************************************

EG11088                  GTGTATGGAGCCGATGTTGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATTCGCG 572
pdhR_7_R.trimmed.seq     GTGTATGGAGCCGATGTTGGCCCAGAATGTCCGCCAGAACTTCGAATTGCTCTATTCGCG 780
                         ************************************************************

EG11088                  TCGCGAGATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTTGAAGCGATTATGGC 632
pdhR_7_R.trimmed.seq     TCGCGAGATGCTGCCGCTGGTGAGTAGTCACCGCACCCGCATATTTGAAGCGATTATGGC 840
                         ************************************************************

EG11088                  CGGTAAGCCGGAAGAAGCGCGCGAAGCATCGCATCGCCATCTGGCCTTTATCGAAGAAAT 692
pdhR_7_R.trimmed.seq     CGGTAAGCCGGAAGAAGCGCGCGAAGCATCGCATCGCCATCTGGCCTTTATCGAAGAAAT 900
                         ************************************************************
```

```
EG11088                 TTTGCTCGACAGAAGTCGTGAAGAGAGCCGCCGTGAGCGTTCTCTGCGTCGTCTGGAGCA 752
pdhR_7_R.trimmed.seq    TTTGCTCGACAGAAGTCGTGAAGAGAGCCGCCGTGAGCGTTCTCTGCGTCGTCTG-AGCA 959
                        ************************************************** ****

EG11088                 ACGAAAGAATTAG-------------------------- 765
pdhR_7_R.trimmed.seq    ACGAAAGAATTGAAGGGCGAATCGAAGCTGGCTTTGCGA 998
                        **********
```

Brief Conclusions: Wed Dec 14, 2005 Finally resequenced the pdhR from above that it looked like it was alright but the read wasn't very good. It looks fine.

1.4 Cloning TOPO constructs into MG1655

Vectors for lrp, rhaR were originally cloned into TOP10 and will be transferred to MG1655 ATCC cells.

Competent cells were made according to the protocol on page 413 section C.2.4. They were digested to check for proper size post-tranformation (see Figure 1.22).

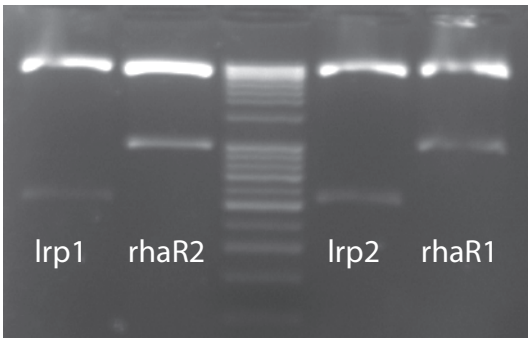


Figure 1.22: 80 ml, 1.5% agarose gel with 0.4 ul of 1% ethidium bromide run for 40 min at 100 volts. 10 μl of Fisher BioReagents exACTGene 1KB Plus DNA Ladder, with bands:weights(ng) of 10000:43, 8000:31, 6000:28, 5000:41, 4000:36, 3000:31, 2500:26, 2000:23, 1500:22, 1000:35, 900:26, 800:52, 700:31, 600:18, 500:35, 400:12, 300:17, 200:12, 100:7 was used. 5μl of the centrifuge minipreps was used.

Mon Dec 5 18:23:28 EST 2005 Additional genes to be transformed into MG1655 are: fecI3, fliA2, nusA1, and ydaK4. 0.5μl of each was added to compentent cells on ice for appx 15min heat-shocked for 30 sec, added 450μl of LB, incubated with shaking at 37 for 45 min. 30μl of each was spread onto an ampicillin plate. nusA1 had no colonies. Others were grown and minipreped to make sure the insert was there.

Also gfp4, lexA5, pdhR7 had been previously miniprepped and eluted into a 96-well tube format provided by qiagen and they all dried out in the fridge. So they were regrown from freezer stocks (prepared right before I did the minipreps) and miniprepped and transformed into MG1655.

1.5 ChIP optimizing parameters

1.5.1 Round1 with lrp:A only

Tue Oct 18 20:37:53 EDT 2005

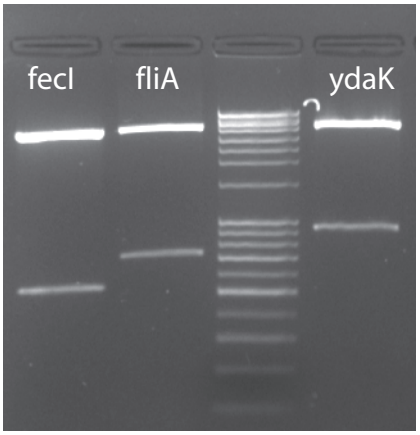


Figure 1.23: 80 ml, 1.5% agarose gel with 0.4 ul of 1% ethidium bromide run for 40 min at 100 volts. 10 μl of Fisher BioReagents exACTGene 1KB Plus DNA Ladder, with bands:weights(ng) of 10000:43, 8000:31, 6000:28, 5000:41, 4000:36, 3000:31, 2500:26, 2000:23, 1500:22, 1000:35, 900:26, 800:52, 700:31, 600:18, 500:35, 400:12, 300:17, 200:12, 100:7 was used. 5μl of the centrifuge minipreps was used.

Trying to figure out best OD to choose and further clarify the link between OD and DNA yield. Also trying to lessen the RNA contamination by using an RNAse cocktail and a pre-phenol:chloroform RNAse digestion for 30min at 37C. All strains are lrp:A, which is the only one that was verified to be error free by sequencing.

Running 4 samples in duplicate (8 total). Procedure begins as in 1.2.1. Two samples R3:A, R3:B were grown for 5hr 15min (R3). The other two samples, R2:A, R2:B, were grown for 3hr 45min (R4).

sample times:
Round3 (R3) in 1PM
Round2 (R2) in 2:30PM
Round1 in 4PM but I broke the flask (there was only going to be one sample of this)
all samples out at 6:15PM

Strain	**growth time**	**OD 600 (-bkgrnd)**	**avg(OD 600)**
R3:A	5hr 15min	0.63, 0.66	0.64
R3:B	5hr 15min	0.59, 0.60	0.59
R2:A	3hr 45min	0.31, 0.32	0.32
R2:B	3hr 45min	0.32, 0.34	0.33

Two 10ml samples were taken from each 50ml sample to make 8 samples: R3:A1, R3:A2, R3:B1, R3:B2, R2:A1, R2:A2, R2:B1, R2:B2. All samples were crosslinked and sheared using the crosslink/shearing protocol in section C.4.1. Sheared lysates were visualized on an agarose and can be seen in Figure 1.24.

Brief Conclusions: Sonication with the digital Branson is much cleaner looking. Some of this might be due to the ability of the newer gel imaging system to make prettier gels. The smears are quite consistent though.

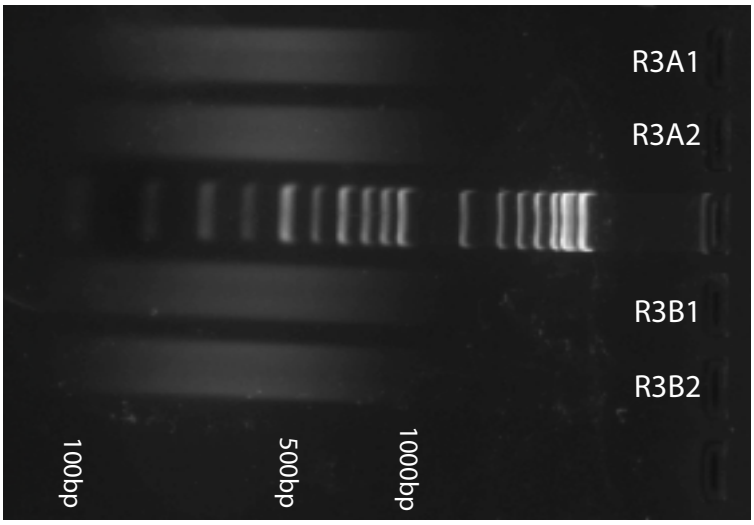


Figure 1.24: 80 ml, 1% agarose gel with 0.5 ul of 1% ethidium bromide run for 40 min at 100 volts. 10 μl of Fisher BioReagents exACTGene 1KB Plus DNA Ladder, with bands:weights(ng) of 10000:43, 8000:31, 6000:28, 5000:41, 4000:36, 3000:31, 2500:26, 2000:23, 1500:22, 1000:35, 900:26, 800:52, 700:31, 600:18, 500:35, 400:12, 300:17, 200:12, 100:7 was used. 3μl of each sheared DNA sample was used. Samples were sonicated using a Branson 250 digital sonifier. R3A1 and R3A2 were sonicated 3x 30sec at 20% power. R3B1 and R3B2 were sonicated 4x 30sec at 20% power.

qPCR To reduce cost I ran only one of the samples (plus one sample with 4 replicates takes an entire plate). The following was for sample R3:A1. The plate was set up to include 4 pcr Replicates.

Reactions were performed in a 384-well Abi qPCR plate. For details see C.4.1. The phenol:chloroform extracted DNA was resuspended in 100μl of 10mM Tris. This was diluted with 302μl of water to ease pipetting into the plate. 40μl was alliquoted into 10 pcr tubes to facilate multichanneling. 4μl of the 400μl total target DNA solution was used per well. This is equivalent to using 1μl of the 100μl resuspension per reaction.

The final qPCR reaction and order of addition to the plate qPCR reaction was:

1. add 12 μl of master mix + water (10μl sybr green master mix, 2μl water) from reservoir
2. add 4 μl of primer mixes (final conc. 800 pM)
3. add 4 μl of dilute target DNA

Out of curiousity, I took a picture of the plate when the qPCR plate finished (see Figure 1.25).

The raw data (Ct values round to the nearest thousandth) can be seen in Table 1.5.1.

The Ct values must be compared to determine a p-value for the enrichment and a α-value must be chosen to determine at what level genes are accepted as having had their promoters positively enriched when using the correct antibody relative to the unrelated antibody. A quick inspection of the no-antibody/beads-only columns reveals that all genes are significant relative to the no-antibody control so it is likely I'll remove this control from my experiments in the future to save time and cash.

Currently, I'm using a one-sided unpaired t-test or a one-sided paired t-test to determine p-value. In the end I should have more replicates than are presented here, so the t-tests will have a little

lrp qPCR plate setup																								
-	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
B	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
C	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
D	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
E	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
F	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
G	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
H	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
I	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
J	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
K	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
L	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
M	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
N	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
O	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-
P	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	blank	-	-	-	-

Table 1.4: lrp:A1 qPCR plate setup. Yellow rows are + control template (sheared lysate DNA). Cyan rows are immunoprecipitated DNA with a specific antibody. Red are immunoprecipitated with an unrelated antibody. White are immunoprecipitated with no antibody.

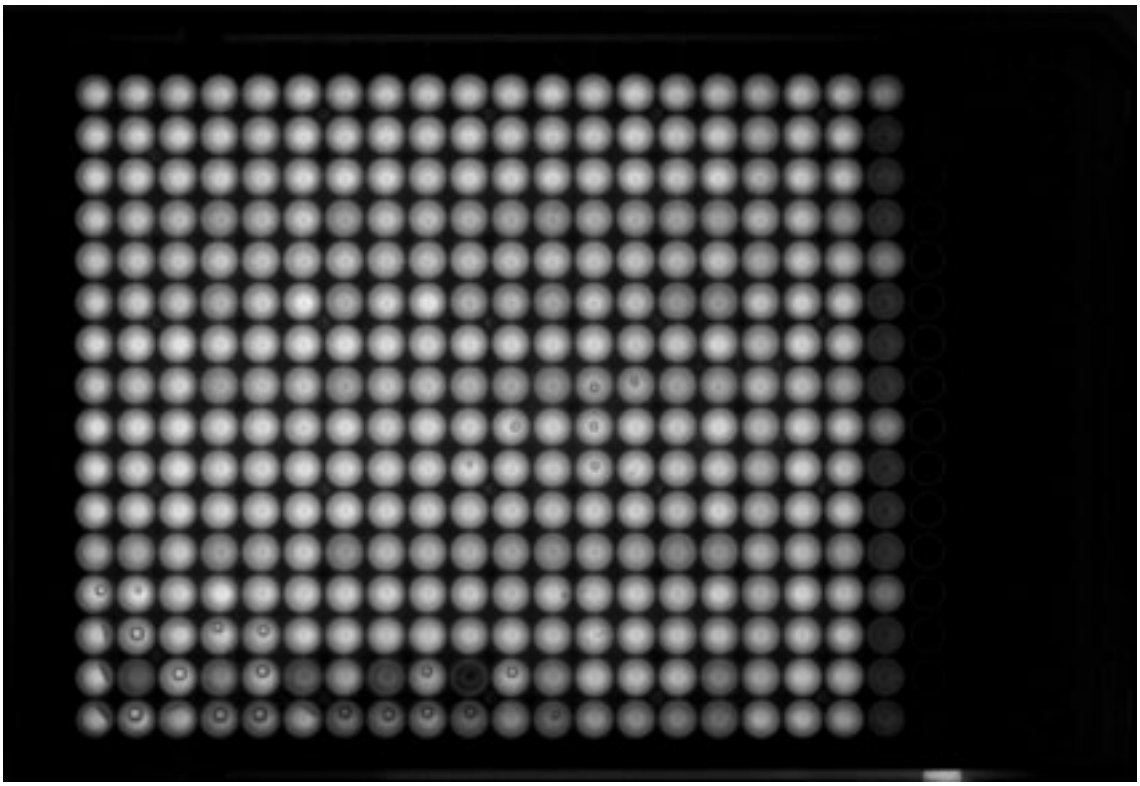


Figure 1.25: qPCR plate from the first round of chipping with lrp. Last 4 columns (black) were water. The little circles are water droplets on the tranparent cover formed when I jiggled the plate a little too much on my way to the imager.

more meaning than a 4 against 4 t-test. I'm thinking 3 sample replicates and 3 qPCR replicates = 9 total replicates will be sufficient.

lrp R3:A3 ChIP results															
-	**pCnt1**	**pCnt2**	**pCnt3**	**pCnt4**	**sAnti1**	**sAnti2**	**sAnti3**	**uAnti1**	**uAnti2**	**uAnti3**	**uAnti4**	**nAnti1**	**nAnti2**	**nAnti3**	**nAnti4**
recA	14.006	14.039	14.011	13.987	25.139	24.945	24.897	25.314	25.479	25.293	25.593	32.077	30.309	30.347	30.390
cysK	14.173	14.283	14.241	15.692	26.478	26.254	26.131	26.918	26.998	26.815	25.585	32.440	32.793	32.148	33.056
entC	13.848	14.019	13.924	13.846	24.245	24.379	24.273	24.421	24.644	24.609	24.761	33.460	35.060	36.558	34.014
dppB	14.511	14.734	14.611	15.306	23.902	23.441	23.211	25.477	25.665	25.383	24.708	-	-	38.019	-
fliF	14.520	15.029	14.640	14.556	26.447	26.713	27.126	26.885	27.448	27.433	27.598	-	-	38.182	-
ilvC	13.609	13.696	15.696	13.637	23.847	23.983	23.791	24.172	24.304	24.213	23.375	36.641	31.985	38.575	35.788
serA	13.792	13.879	13.821	13.813	23.114	23.009	23.279	24.714	24.547	24.578	24.560	38.939	-	38.414	-
lysC	14.000	14.022	14.215	13.984	24.887	24.538	24.662	25.280	25.270	25.250	25.184	32.378	30.427	32.859	-
livK	13.704	13.956	13.768	13.770	20.212	20.159	20.233	23.743	23.963	23.509	23.861	36.320	38.059	39.246	-
metA	13.651	13.651	13.668	13.654	23.905	23.863	23.861	24.278	24.273	23.775	26.687	28.233	30.776	29.501	-
aroG	13.931	14.055	13.969	14.198	24.281	23.223	23.529	24.573	24.722	25.003	24.657	-	-	-	-
metE	14.510	14.440	14.444	14.501	29.197	29.213	29.461	29.382	29.610	29.448	30.338	37.004	-	-	-
aroL	13.941	14.046	13.957	13.937	25.218	24.312	24.833	24.690	24.789	24.582	24.915	-	-	38.248	-
metF	13.968	14.219	13.943	13.941	24.622	24.448	24.799	24.915	24.951	25.106	25.130	-	-	39.012	-
aroP	13.939	14.037	13.972	14.007	24.156	23.905	24.146	24.433	24.417	24.314	24.547	35.690	36.375	-	-
dapB	13.790	13.732	13.754	13.870	23.999	23.869	23.777	24.618	24.518	24.671	23.924	31.301	32.195	33.531	38.918
codB	-	-	-	-	-	39.360	-	-	-	-	39.936	-	-	-	-
dapD	14.076	14.102	14.130	14.184	24.637	24.561	24.245	24.575	24.694	24.579	24.863	38.938	-	-	36.849
cysC	14.662	14.803	14.821	15.070	16.640	16.873	16.227	16.573	17.005	16.727	16.780	19.960	20.215	19.621	19.222

Table 1.5: qPCR Ct values rounded to the nearest 1000th. pCnt = positive control, sheared lysate not precipitated; sAnti = target DNA, immunoprecipitated with the antibody specific to the TF of interest; uAnti = target DNA, immunoprecipitated with an unrelated antibody (myc for an XPress tagged protein), nAnti = beads only, no antibody control. All samples except the sAnti have four replicates. One of the sAnti replicates was lost by pipetting into the wrong hole.

gene	**paired t-test**	**unpaired t-test**
recA	0.0586	0.0047
cysK	0.3537	0.2513
entC	0.0449	0.0092
dppB	0.0033	9.3641×10^{-4}
fliF	0.0129	0.0335
ilvC	0.4308	0.3038
serA	0.0023	4.4390×10^{-6}
lysC	0.0140	8.0723×10^{-4}
livK	2.6981×10^{-4}	3.5336×10^{-7}
metA	0.1852	0.1548
aroG	0.0658	0.0068
metE	0.0959	0.0970
aroL	0.5845	0.5697
metF	0.0332	0.0064
aroP	0.0068	0.0050
dapB	0.0577	0.0238
codB	NA	NA
dapD	0.4970	0.0928
cysC	0.3302	0.1809

From the above results it looks like with an unpaired t-test serA, livK, lysC, dppB, and aroP are significantly enriched. For the paired it appears that aroP would not be included. This interpretation is taken by looking at genes included because they were not likely lrp targets (i.e. negative control genes): recA, entC, fliF and looking at their p-values. You can see that there is often some enrichment even for these negative control genes, so we must be careful with our thresholds. Hopefully, adding more samples will tighten the p-values. The positive controls (known targets serA and livK) are very enriched for both types of t-test which is promising.

Going to repeat the qPCR reaction using R3:A2

Brief Conclusions: Used 2ul of mg/ml glycogen and non-siliconized tube: much easier to see pellet. For the master mix from the reservoir, I added enough for around 5 extra reactions to

account for the lose in the reservoir, this was definitely not enough, I was about 10 wells short. Next time I should probably dilute the target DNA less (e.g. use 3μl instead of 4μl). To allow a little more water to be added to the reservoir. Also, should add extra for around 10 reactions (sucks cause this is expensive).

OD vs DNA yield

One monoclonal vs two monoclonal

To preclear or not to preclear

This step was not tested.

1.5.2 Round1 with lrp:A2

Thur Nov 10 17:47EST, 2005

A quick note, these tables showing where the genes are on the chip are actually wrong. They derive from the table on page 29 that shows the primer layout. Using that table you see all of the qPCR tables thus far presented show rows D and E concatenated to make 24-wells. But in a qPCR 384-well plate, the multichannel pipettor output spacing is such that only every-other well is filled by the pipettor. Thus two pipettes worth, slightly shifted make the entire 24-wells fill up. The tables should show D1,E1,D2,E2,etc... Staggering the two columns.

lrp qPCR plate setup: round 2																							
-	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
B	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
C	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
D	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
E	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
F	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
G	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
H	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
I	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
J	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
K	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
L	recA	entC	fliF	serA	livK	aroG	aroL	aroP	codB	cysC	cysK	dppB	ilvC	lysC	metA	metE	metF	dapB	dapD	5kbUP	1kbUP	pntA	gltB	stpA
M	blank	blank	blank	blank											empty									
N												empty												
O												empty												
P												empty												

Table 1.6: lrp:A1 qPCR plate setup. Yellow rows are + control template (sheared lysate DNA). Cyan rows are immunoprecipitated DNA with a specific antibody. Red are immunoprecipitated with an unrelated antibody. White are immunoprecipitated with no antibody. Blanks are template + master mix only, no primers. 5kUP and 1kbUP were added to see how good the shearing is. pntA is a new potential target from Boris' algorithm. gltB and stpA are new extra positive controls.

1.5.3 Round1 with lrp:B1

Fri Nov 11 17:53:52 EST 2005

This is one of the two samples that used both His and Xpress antibody. In addition these samples underwent a four shearings instead of three. The layout is identical to that from lrp:A2 except only 1.5μl of primer was used (300nM) instead of the normal 800nM, and there were not empty wells with primer only.

1.6 lrp ChIP in Minimal and Rich media in MG1655

In the first round of ChIP their was definitely enrichment for the positive controls, but not that many of the predictions were verified (around 5-9 I think). It was also noisy as everything enriched including the negative controls, so determining the true interactions was done by looking at things enriched compared with the negative controls. It appears that just too much lrp is binding *everywhere*.

This round I'm moving from 1mM IPTG to 100 μM. In addition all cultures have 0.5% glucose which should slow expression from this lac repressed promoter (according to the pTrcHis Invitrogen manual). The minimal media is also the condition where lrp is doing stuff, and hopefully a more diverse expression of different transcription factors will lower the background. Last, the plasmids were cloned in to MG1655 rather than the original TOP10 cloning strain. All samples were grown as 50ml cultures in 250ml baffled flasks with shaking at 300rpm at 37C.

Davis minimal media was used because it is easy to prepare and *very* minimal.

1.6.1 growing lysing shearing

First round on Saturday failed as I didn't realize how slow the culture would grow in minimal media when they have a vector.

Sat Nov 19 around 4PM inoculated two cultures lrp:J and lrp:K freezer stock into Davis minimal with 1% glucose

Sun Nov 20 19:40:00 EST 2005 inoculated 3 250ml baffled flasks with 1:100 dilution of lrp:K (OD 1.04)

Mon Nov 21 09:15ish overnite was too long :(.....

Mon Nov 21 09:30ish started growing from 1:100 dilution into another baffled flask

Mon Nov 21 15:24:44 EST 2005 incubated the LB/1% glucose 1:100 from overnite will grow both. At this time I also added the IPTG.

Mon Nov 21 5:29PM 2005 removed samples from incubator because minimal samples were getting overgrown

Two samples were taken from the initial 6 samples (lrpMa, lrpMb, lrpMc, lrpOa, lrpOb, lrpOc) and OD600 absorbances were take:

Strain	OD 600
lrpMa	0.693
lrpMb	0.726
lrpMc	0.736
lrpOa	0.290
lrpOb	0.283
lrpOc	0.297

It's clear that the minimal strains (lrpM*) were grown more than the LB cultures (lrpO*), but I didn't want one of the cultures to have a lot more time in the prescence of IPTG, and thus much higher Lrp-tagged protein concentration. The twelve total samples were lysed and sheared according to the protocol starting on page 417.

The concentration of each sample was checked on the nanodrop:

Sample	DNA (ng/μl)	260/280	260/230	μl for 25 μg	μl dilution buffer
lrpMa1	217.0	1.97	2.04	115.2	1037
lrpMa2	240.0	1.94	1.99	104.2	938
lrpMb1	241.0	1.97	2.17	103.7	933
lrpMb2	228.0	1.94	2.07	109.6	986
lrpMc1	238.7	1.94	2.05	104.7	942
lrpMc2	255.2	1.96	2.07	98.0	882
lrpOa1	202.6	1.99	2.19	123.4	1111
lrpOa2	180.8	1.96	2.01	138.3	1225
lrpOb1	188.2	1.99	2.14	132.8	1195
lrpOb2	211.7	1.99	2.12	118.1	1063
lrpOc1	216.4	1.98	2.13	115.5	1040
lrpOc2	164.2	1.97	2.14	152.3	1371

Approximately 600ng of all twelve samples (3μl) was loaded onto an agarose gel (see Figure 1.26).

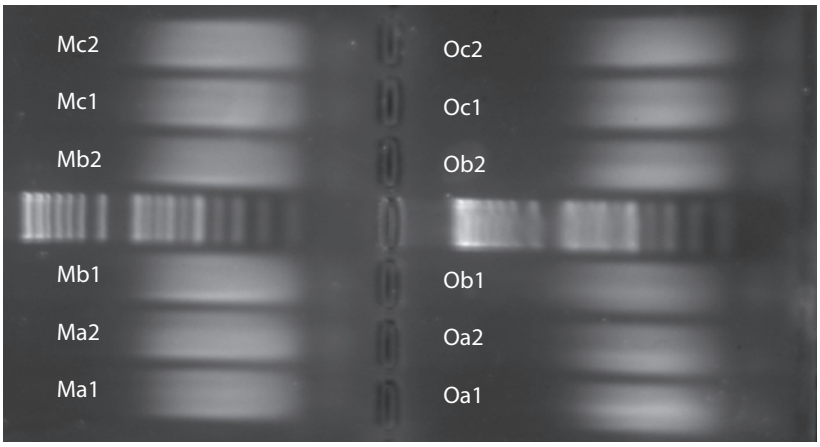


Figure 1.26: 80 ml, 1.5% agarose gel run for 40 min at 120 volts, poststained in 100 ml of H_2O with 0.5 ug/ml ethidium bromide for 40 min followed by 20 min in water alone (both on orbital shaker at 50rpm) to reduce background. Shearing range appears between 1000 and 100 basepairs.

Brief Conclusions: Shearing range is beautiful; DNA yields are a little smaller than I'd like. I'm surprised how similar the LB and minimal media yields are given the OD was 2x higher for the minimal media. However, when spinning down the minimal not all cells were pelleted (the pellet was still much larger than for the LB cultures). I'll proceed to immunoprecipitation after the Turkey break...

1.6.2 Misc.

Negative control primers are going to be rechoosen randomly this time rather than the previous 5 that were chosen by picking targets of other proteins I'm testing. I choose the numbers by picking 20 random genes from the 4345 genes in the microarray and chosing the first 12 in order that were more than 5000bp from a predicted target of all of the TFs I'm testing and that were not hypothetical genes.

1.6.3 Bead washing

Thu Dec 1 21:55:48 EST 2005 precleared samples set to rotate Fri Dec 2 21:52:02 EST 2005 put washed complexes on heat block overnite to reverse crosslinks

I ran the 36 samples in two batches of 18. This is a LOT of work takes 5 hrs a batch. Makes for a very unpleasant day. During the breaks I did more research on how I can make this work in a 96-well format (written in section 1.6.4

The 12 original samples (each with A, B, C for correct antibody, incorrect antibody, and bead only control) were split into the following random blocks of six (with the A,B,C kept together for 18 samples in each block. The samples in the two sets where pick randomly to be: 4, 9, 5, 7, 3, 8; 6, 2, 10, 11, 1, 12. The numbers correspond to their placement in the DNA yield table above (i.e. 4 = lrpMb2, 9 = lrpOb1, etc. . .).

1.6.4 Ideas for higher-throughput and automation

Here's what I learned (most of this I already planned earlier).

The initial growth part is tricky. I grow 50 ml, take a 15 ml sample and end up with around 200μg of DNA, I need 50μg if I assume only one of the negative controls is important. Really 55μg would be nice so I can have a good positive control. With roughly 13.33μg DNA per ml I need around 5ml to make this work. Then I'd need to scale down the lysing procedure accordingly. 2 ml is the biggest 96-well block I know of. Perhaps they go up to 5 ml, but my BioHit pipettor doesn't....

(Small addition: Tue Jan 17 12:00:20 EST 2006 there are 48-well plates that use 5ml of samples; that's probably the perfect size. 48-cultures can be split into two correct antibody/incorrect antibody. The resulting 96 (i.e. 48 sonicated lysates split into two) will go into the 96-well plates. If that's not enough culture, there are 24-well 10ml plates, but that would involve setting up and distributing from many different plates.

Also would be very useful to buy a thermo-sealer AB-0384 $1500; need this too AB0724 a completely automated one costs 30K ouch! AB-0950)

For the shearing, it could be done with one of these bioruptors if this machine worked. http://www.diagenode.com/I Costs around 12 grand. If it worked it would really cut down on the worst step. Machine is not set up for 96-well, but even as it is doing 12 samples in falcon tubes in 10 minutes with out my having to open and hold each tube brings tears to my eyes.

bead washing with 96-well costar filtrex plates: much faster! especially if I had a robot to load the 96-wells!

qPCR robots Corbett is cool but only has 96-well

interesting unrelated robot thing: genomic solutions has cool way to grow LOTS of cells biotek robot doesn't do what I want neither does this one ttp labtech

Lissy 2002 and bu has one cmld.bu.edu/instrumentation/lissy.html problem is it doesn't go that low in volume

the hummingbird looks just as good as the deerac looks good to but it is slow

deerac has by far the best one, but it only accommadates one extra plate to transfer to the production plate. Not sure if it can do things like A12 sample to H23 production. Wouldn't be too bad if it can handle 384 in both places. Equator HTS or GX pipettes down to 50nl!!!!! Smaller qPCR reactions could save a lot of dough. Plus doesn't need tips.

(Thu Jan 19 13:40:06 EST 2006 labchip 90 would be nice for operon prediction, we could run normal pcr's in 384 well plates, only need 150nl. Gives list of bands, so we could multiplex at different band sizes and get a moderately quantitative reading (certainly good enough for a yes, no answer). 384-well plates can resolve between a 140bp and a 210bp fragment. Also between a 400 and a 420bp fragment. So we might get 5-10 rxns per well. Total pcr rxns is 8690 (17380 primers $$$), could be done in 3, 384-well plates. Would take 2hrs for the PCR and 12 hrs for the subsequent gels. One gel can run 1800 rxns, so you could check the genomes operons 2x per gel.

1.6.5 qPCR

new primers were designed this time we have 12 random genes (rather than just genes from other targets that I guessed wouldn't be bound by the TF of interest). The random genes were picked using a few rules.

1. generate a random list of genes (did in matlab from all 4345 genes I pick 150 or so).
2. throw away genes that are putative (to be far we want to compare apples and apples, so we should make sure the protein is most likely a real one
3. throw away things that are in the list of the top 40 targets or are within 5kb of one of those genes (by looking at the ecocyc genome browser)

Also, primers are now ordered in 96-well plates from IDT prenormalized to 100 μM. Primer pairs are placed consecutively by column (e.g. if the forward primer is in well C3 the reverse will be in D3), to make for creation of primer mixes. It's also cheaper and MUCH faster for me not to have to keep track of all that stuff now.

From now until further notice, all TFs will have their top 36 interactions tested (minus a few positive controls). Combined with the 12 random genes that's 48 targets per TF times the number of replicates (4-6 haven't decided yet) times the number of controls (4) for a total of 768-1152 qPCR rxns per TF. Yikes that's a lot. If I ever do this again I need to drop one of those controls as it'll reduce the number of rxns by a forth.

New list of lrp targets

The previous 36 targets have been altered. To be scientifically honest, I felt I shouldn't just remove a bunch of stuff to fill it with Boris' algorithms predictions and try again, especially since only two genes appeared to be new targets. I only removed one predicted target codB from the previous plate because those primers never worked, so I can say nothing about that gene plus or minus. Also, one new positive control serC was added.

The major change however was for the negative control genes. In order to get a better estimate of the background noise a true set of random genes was chosen. This time we have 12 genes rather than the previous 3-5 and they aren't just chosen from the list of primers I already had. All of the previous negative control primers have been removed. The 1kb and 5kb controls are still there, but will likely not be used for other TFs I check.

Specifically the genes removed were codB, recA, fliF, and entC. New targets are shown in italics in the new primer table (see Table 1.6.5).

updated lrp target plate												
-	1	2	3	4	5	6	7	8	9	10	11	12
A	metN	thrL	yagU	cvpA	serC	yhjE	leuLABCD	gdhA	purC	argI	argA	yebR
B	argG	nlp	ompT	purM	ilvC	metE	metF	metA	serA	livK	lysC	pntA
C	dppB	stpA	aroG	cysK	aroL	cysC	aroP	gltB	dapD	dapB	5kb	1kb
D	gcl	mog	pinO	idnD	yhaF	nhaA	amiA	goaG	kdtB	yagG	citC	fruK

Testing the primer plate The new primer plate needs to be tested to make sure all the primers work. Unfortunately, I'm having issues with this simple task.

Yesterday, Thur Dec 8 2005, I tried running the 48 PCRs in a plate rather than in PCR strips. No luck. I used a Costar Thermowell 96-well plate (part number 6551), which has the nice feature that it's easy to break it into sections (e.g. to use in our PCR machine which only has 48-well blocks. I sealed the plate with Costart Thermowell Sealers (part number 6524). Unfortunately, getting a good seal was not easy, especially since the plate was broken in half, leaving a proper edge on only one side. The bad seal led to the evaporation of most of the solution in the plate toward the bottom right. As you went towards the top left, it didn't look as bad. The results show that the bottom of the plate didn't work well. But also for some reason (which I don't know) the top of the plate didn't work either. Only the middle really worked (see Figure 1.27.

Valuable Lesson: Make sure to seal your 96-well PCR plates well to avoid evaporation.

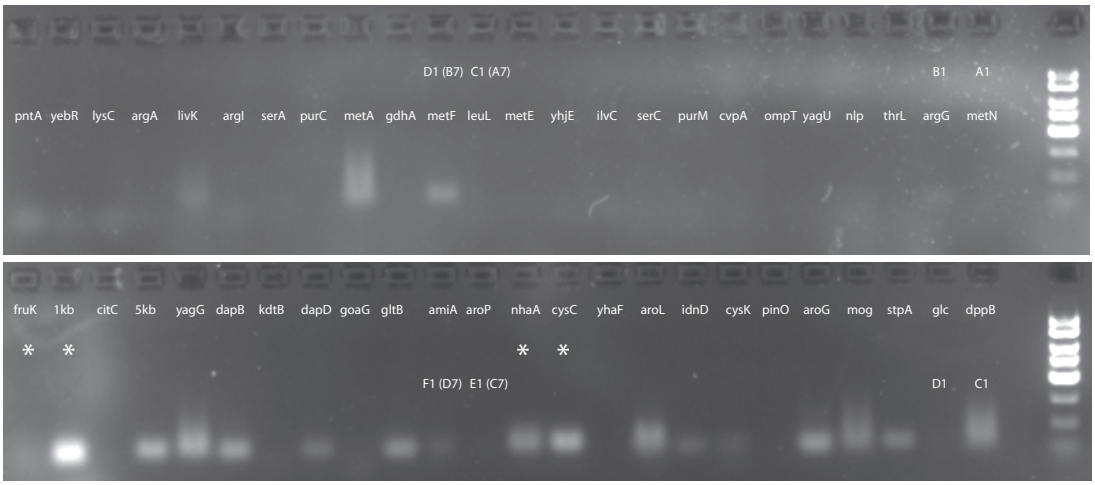


Figure 1.27:

Then this morning, I ran a new PCR dropping the volume down to 25 μl (previous was my typical 50 μl). However, before loading the gel, I realized I forgot to add the template DNA to the mix. Now I'm rerunning the 25 μl reaction (I lowered the volume for these reactions for two reasons: 1) why not, I don't need much DNA and this will be cheaper; 2) I'm very low on Taq, in fact I ran out of the cheap Qiagen Taq, and this round I'm using the Easy-A proofreading Taq.

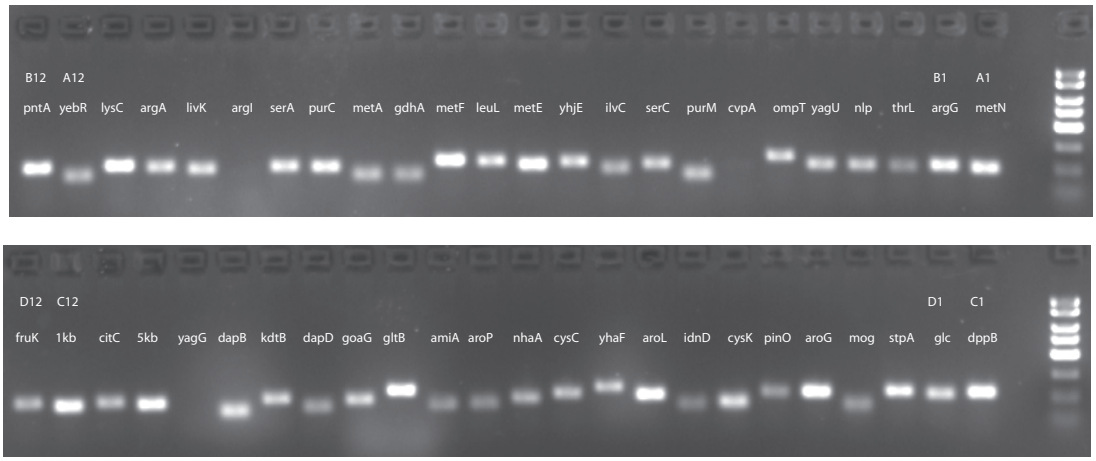


Figure 1.28: 300 ml, 1.5% agarose gel with 4.5 ul of 1% ethidium bromide run for 40 min at 100 volts. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used. 8μl of the PCR/dye was loaded per lane.

Brief Conclusions: It looks like the primers work fine when I don't screw up the PCR (see Figure 1.28. The only problem is that three reactions didn't work. I'll know better when I run the qPCR if they were just random failed PCRs or bad primers. Since all previously verified primer pairs worked I think it's probably bad primers. I don't mind some failures, as I'm planning to test 36 per gene and dropping a couple doesn't hurt too much. Replacing them is unweildy given the way I pipette these by hand with a multchannel, with a robot, it'd be easy to replace them, but that's another day. The only problem is that one of the failures is a random gene, which knocks my number of negative control genes for *every* TF I test down to eleven, which is an odd number to have and weakens our error model. All in all I'm happy with the new plate oligo format from IDT and I need to hurry up and order the other plates.

The layout for the 384-well qPCR plate derives from the primer place and is shown in Table 1.6.5.

Lrp minimal vs rich qPCR setup																								
-	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	metN	argG	thrL	nlp	yagU	ompT	cvpA	purM	serC	ilvC	yhjE	metE	leuLABCD	metF	gdhA	metA	purC	serA	argI	livK	argA	lysC	yebR	pntA
B	dppB	gcl	stpA	mog	aroG	pinO	cysK	idnD	aroL	yhaF	cysC	nhaA	aroP	amiA	gltB	goaG	dapD	kdtB	dapB	yagG	5kb	citC	1kb	fruK
C	metN	argG	thrL	nlp	yagU	ompT	cvpA	purM	serC	ilvC	yhjE	metE	leuLABCD	metF	gdhA	metA	purC	serA	argI	livK	argA	lysC	yebR	pntA
D	dppB	gcl	stpA	mog	aroG	pinO	cysK	idnD	aroL	yhaF	cysC	nhaA	aroP	amiA	gltB	goaG	dapD	kdtB	dapB	yagG	5kb	citC	1kb	fruK
E	metN	argG	thrL	nlp	yagU	ompT	cvpA	purM	serC	ilvC	yhjE	metE	leuLABCD	metF	gdhA	metA	purC	serA	argI	livK	argA	lysC	yebR	pntA
F	dppB	gcl	stpA	mog	aroG	pinO	cysK	idnD	aroL	yhaF	cysC	nhaA	aroP	amiA	gltB	goaG	dapD	kdtB	dapB	yagG	5kb	citC	1kb	fruK
G	metN	argG	thrL	nlp	yagU	ompT	cvpA	purM	serC	ilvC	yhjE	metE	leuLABCD	metF	gdhA	metA	purC	serA	argI	livK	argA	lysC	yebR	pntA
H	dppB	gcl	stpA	mog	aroG	pinO	cysK	idnD	aroL	yhaF	cysC	nhaA	aroP	amiA	gltB	goaG	dapD	kdtB	dapB	yagG	5kb	citC	1kb	fruK
I	metN	argG	thrL	nlp	yagU	ompT	cvpA	purM	serC	ilvC	yhjE	metE	leuLABCD	metF	gdhA	metA	purC	serA	argI	livK	argA	lysC	yebR	pntA
J	dppB	gcl	stpA	mog	aroG	pinO	cysK	idnD	aroL	yhaF	cysC	nhaA	aroP	amiA	gltB	goaG	dapD	kdtB	dapB	yagG	5kb	citC	1kb	fruK
K	metN	argG	thrL	nlp	yagU	ompT	cvpA	purM	serC	ilvC	yhjE	metE	leuLABCD	metF	gdhA	metA	purC	serA	argI	livK	argA	lysC	yebR	pntA
L	dppB	gcl	stpA	mog	aroG	pinO	cysK	idnD	aroL	yhaF	cysC	nhaA	aroP	amiA	gltB	goaG	dapD	kdtB	dapB	yagG	5kb	citC	1kb	fruK
M	metN	argG	thrL	nlp	yagU	ompT	cvpA	purM	serC	ilvC	yhjE	metE	leuLABCD	metF	gdhA	metA	purC	serA	argI	livK	argA	lysC	yebR	pntA
N	dppB	gcl	stpA	mog	aroG	pinO	cysK	idnD	aroL	yhaF	cysC	nhaA	aroP	amiA	gltB	goaG	dapD	kdtB	dapB	yagG	5kb	citC	1kb	fruK
O	metN	argG	thrL	nlp	yagU	ompT	cvpA	purM	serC	ilvC	yhjE	metE	leuLABCD	metF	gdhA	metA	purC	serA	argI	livK	argA	lysC	yebR	pntA
P	dppB	gcl	stpA	mog	aroG	pinO	cysK	idnD	aroL	yhaF	cysC	nhaA	aroP	amiA	gltB	goaG	dapD	kdtB	dapB	yagG	5kb	citC	1kb	fruK

Table 1.7: yellow is positive control, cyan = A (correct antibody), magenta = B (incorrect antibody), white = C (beads only)

The schema in Table 1.6.5 allows two replicates per plate (hence the duplicated 1970's rainbow pattern). Takes six plates to run 12 chIP samples. Yikes, I'd better buy more qPCR master mix!

Each well will be filled as follows:

1. add 15.5 μl master mix + water
2. add 1.5 μl primer mix (150nM)
3. 3 μl of template

Master mixes are prepared for 8 extra reactions. qPCR master mix (3940μl) + water (2167). For template add 100 μl of water to each tube (200μl total), allow for appx 65 rxn, we're running 48.

I draw a random number to choose which sample (O or M) should occupy the top half of the plate (in case there are differences in the machine across the plate).

First batch plate A302JWOB Sun Dec 10, 2005

M first then O

Ran short by 7 wells of master-mix + water. Pipetted those by hand (P12, P14, P16, etc). One other slight problem, when I spin down the plate prior to walking it over to the LSEB building where the 7900HT is, the centrifuge was at 4C. Normally this step puts all the sample on the bottom of the wells, but the temperature difference made a little condensation form on the top of each well (a very small amount but still annoying). I let the lid of the centrifuge stay open for a while and it pretty much fixed the problem.

Valuable Lesson: Warm the centrifuge to room temperature before attempting to spin down a room temp qPCR plate.

No other mishaps occured to my knowledge. Rows A-H contain sample lrpMa1. Rows I-P contain sample lrpOa1. Just a reminder M = Davis + 0.5% glucose, O = LB + 0.5% glucose. The batch took a little less than 1hr 30min to set up. Remaining templates were put back in freezer.

Second batch plate A302JWOA Sun Dec 11 16:00 EST 2005

O first then M

Ran way short of master mix + water. Had to make more for entire last row.

Third batch plate A302JWO9 Sun Dec 11 18:00 EST 2005

M first then O

Figured out how to stop running out of master mix + water. When pipetting the 15.5 μl from the multichannel I would press until the pipette hit the firm part and the press a little further to get the last little bit out. This has the negative affect of creating bubbles. It seems it also makes you run out of material faster. I wouldn't have guessed that. I need to repeat this and see if that's true, but I didn't run out of reagent this time.

Valuable Lesson: only aspirate the multichannel until the first stopping point when pipetting master mix + water

This round was the most error prone so far. It was a little from fatigue (768 PCR rxns in one day is a little much), but mostly because the silicon grease must have worn off in the two previous batches. From now on I need to grease the low volume multichannel before every qPCR plate. There were a number of times the tips didn't stick to the channel. One time I ended up pipetting 300nM into 12 wells (F1,F3,F5,...,F23), because the first time not all the tips sucked up the 1.5μl and I lost track of which ones didn't. Hopefully, primer concentration doesn't have a huge impact, we'll have to take a closer look at that row.

Valuable Lesson: Apply silcon grease to the low volume multichannel pipettor before each round of qPCR or you'll wish you had

Fourth batch plate A302JWO8 Mon Dec 12 20:54:08 EST 2005

M first then O (I seem to be good at generating a random number less than 0.5 (which means I do M first)

No problems that I saw; I'm becoming a good robot.

Fifth batch plate A302JWNS Mon Dec 19 13:28:22 EST 2005

M first then O.

No problems that I saw.

Sixth and final batch plate A302JWNT Mon Dec 19 21:44:22 EST 2005

O first then M.

No problems that I saw.

1.7 ChIPrndC: pdhR, fecI, lexA, ydaK

Wed Dec 7 11:23:33 EST 2005 cells in incubator in davis with 0.5% glucose for overnite growth (which takes 24hrs to hit stationary with minimal)

I thought they didn't grow, but they were just growing *very* slow.

ydaK is getting the can. While designing the primers, I noticed that it's predicted targets are all right next to each other. Typically this is a good sign as neighboring regions of bacterial genomes are often regulated by the same TF. Then I noticed fnr which seemed fishy and was also a neighbor, and I saw the problem. Almost half of the chips were run on the strain Jamey got from Tim Gardner, which I noticed had an fnr deletion (Jamey then found a paper describing this deletion in the Yale stock that is NOT in the stock used by Blattner to sequence the genome). These neigboring genes probably appear corregulated because they are all correlated due to the knockout...

Tue Dec 13 22:36:33 EST 2005 Finished with pdhR and lexA; fecI didn't make it, grew too slow. I would've guessed pdhR would've been the slow one. Trying fecI again tomorrow. Did minipreps to make sure tubes were labeled correctly (they were; see Figure **??**. Also saved 1ml of each sample to try SDS-page with his-stain to check protein expression.

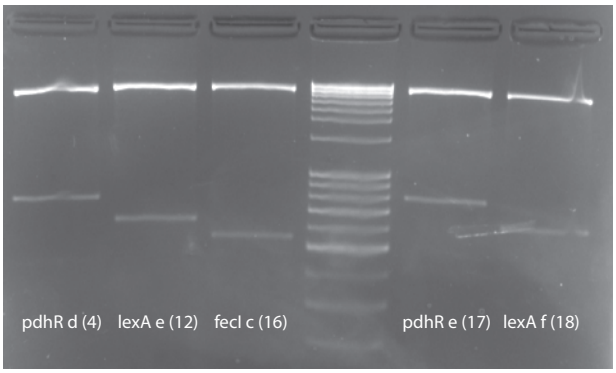


Figure 1.29:

Wed Dec 14, 2005 Finished growing lysing, shearing fecI. Minipreps from previous day checked out ok; seems to be no problems with mislabeling fecI just grows slow. This time fecI was started *without*; it seemed to grow much faster. 100mM IPTG was added 2hrs prior to adding formaldehyde (unfortunately an hour less than the other two samples). Also the cell grew a bit more dense than I would've liked (but I was trying to increase exposure to IPTG).

Sample	**OD600**	**growth time**	**DNA (ng/ul)**	**260/280**	**260/230**	**μl in 25 μg**	**μl dilution buffer**
fecI F	0.680	5 hr 2 min	376.6	1.99	2.15	66.4	597
pdhR B	0.592	3 hr 20 min	407.3	1.99	2.09	61.4	552
fecE E	0.727	5 hr 2 min	459.9	1.99	2.18	54.4	489
pdhR D	0.658	3 hr 20 min	468.4	1.98	2.14	53.4	480
pdhR A	0.657	3 hr 20 min	422.5	1.98	2.09	59.2	532
fecI A	0.710	3 hr 20 min	382.0	2.00	2.19	65.4	589
pdhR C	0.658	3 hr 20 min	444.4	1.98	2.15	56.3	506
lexA D	0.317	3 hr 20 min	173.3	2.00	2.15	144.3	1298
pdhR F	0.635	3 hr 20 min	413.1	1.98	2.10	60.5	545
lexA C	0.273	3 hr 20 min	137.9	2.02	2.19	181.3	1632
lexA B	0.260	3 hr 20 min	123.7	1.99	2.16	202.1	1819 (1738)*
lexA E	0.272	3 hr 20 min	137.4	2.00	2.16	182.0	1638
lexA A	0.269	3 hr 20 min	111.5	2.00	2.14	224.2	2018 (1716)*
fecI B	0.701	5 hr 2 min	462.8	1.99	2.16	54.0	486
fecI D	0.694	5 hr 2 min	434.5	1.99	2.18	57.5	518
fecI C	0.689	5 hr 2 min	392.2	1.98	2.09	63.7	574
pdhR E	0.671	3 hr 20 min	403.0	1.99	2.22	62.0	558
lexA F	0.271	3 hr 20 min	128.5	1.99	2.17	194.6	1751

The shearing range was checked by running appx 600ng on a 1.5% agarose gel. Unfortunately the gel pour wasn't the best, I ran the gel too long and my samples evaporated a little on the parafilm (I know when one thing goes wrong, they all do!), so it's not the most beautiful gel (see Figure 1.30). But if you look close, you see the shearing range is the typical one or perhaps a little shorter (I think the shorter is just because there is less DNA in the gel than I typcically run), but overall looks good as is typical for this step now.

Thu Dec 15 23:13:46 EST 2005 Just added the antibodies. ChIPing sucks some times. I gotta find a way to shorten this thing into 24-48 samples a day.

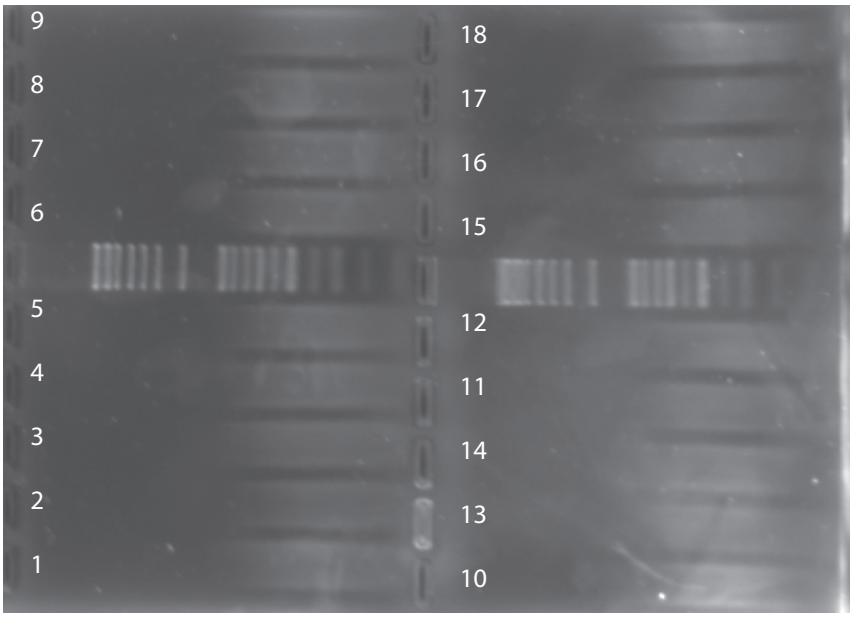


Figure 1.30: Sheared DNA samples were run on a 1.5% gel to check the size range, which appears to be in the range of 100-800bp.

1.7.1 primer design

Wed Dec 7 16:53:58 EST 2005

Primers are now designed automatically with a couple scripts I wrote that are now in the CVS (numerical/primer_design) and the publically available primer3 software. Primer order is randomized except all 12 random genes need to be kept together to simplify pipetting. However, the row which the random genes are placed in (A-D) is chosen randomly.

lexA primer plate, random genes are in row A												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	gcl	mog	pinO	idnD	yhaF	nhaA	amiA	goaG	kdtB	yagG	citC	fruK
B	lasT	yfiR	galE	araJ	cspG	ydjM	araA	dcrB	araE	sufA	ydaN	fnr
C	aceE	fldB	cytR	nhaR	polB	recN	b1141	dinB	mdh	ydaM	dinD	b1458
D	cspI	envY	aceA	galP	dinG	fliA	oraA	dinF	hemC	dinI	dnaA	rfaB

fecI primer plate, random genes are in row D												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	yedF	yebG	wzxE	wecD	sieB	yrdC	yncE	cspG	mutT	ddlB	entF	map
B	bfd	fecA	amiB	exbB	fdhE	feoA	rffG	fecI	fhuF	wecG	ynaI	mntH
C	entA	sodA	proV	iscA	oraA	fhuA	nrdE	apaG	dinF	sulA	ydiE	ybaN
D	gcl	mog	pinO	idnD	yhaF	nhaA	amiA	goaG	kdtB	yagG	citC	fruK

Primers were tested by PCR.

1.7.2 qPCR

It will take 9 full 384-well plates to complete this experiment. Two samples will be run in each plate as in Table 1.6.5, except the samples will be run in the order of the initial randomization

pdhR primer plate, random genes are in row D												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	b3171	fdhE	cspG	yjaG	map	yrdD	yjgP	yfiR	rrmB	ndh	cdsA	ybeB
B	yjeQ	aceE	ispB	ubiA	ybjM	dinF	oraA	rrnG	ubiB	ubiD	alr	yhbE
C	lpxA	rfaB	dinG	ybaP	ddlB	yebG	yrbF	b3790	cspB	yhhF	fecI	fecA
D	gcl	mog	pinO	idnD	yhaF	nhaA	amiA	goaG	kdtB	yagG	citC	fruK

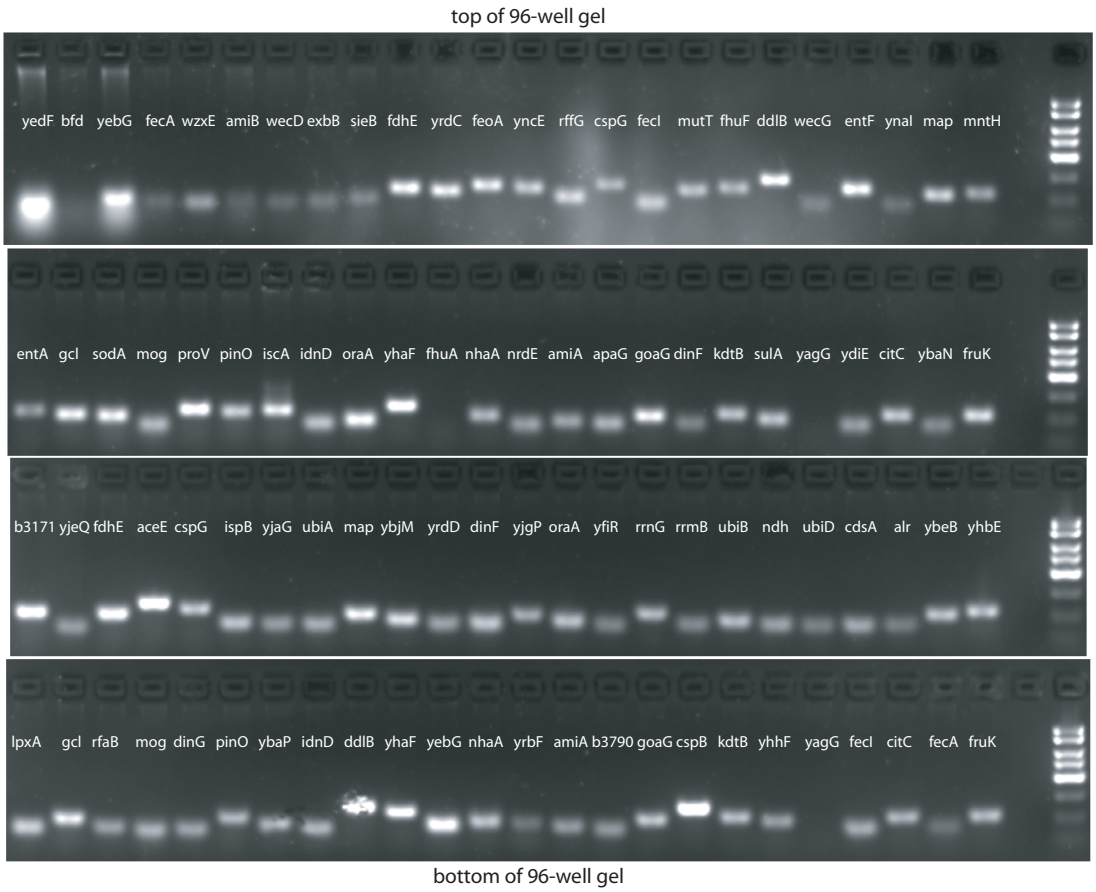


Figure 1.31: 300 ml, 1.5% agarose gel with 1.8 ul of 1% ethidium bromide run for 40 min at 100 volts. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used. 8μl of the PCR/dye was loaded per lane. The strange thing on the fourth row a kind of fuzzy aberation in the image is a piece of scotch tape that feel into the gel. Fortunately it didn't seem to affect the way gel ran. The first two rows are fecI; the last two are pdhR.

(the order is the one used in the table above showing the sheared DNA yields for each sample). For the previous ChIP experiment each plate contained one lrpMinimal sample and one lrpLB sample. These plates may contain samples from different TFs as I'll just go down the random list consecutively.

The target ordering can be inferred be looking at the individual 96-well oligo plates detailed in the tables above and combining two rows with alternating columns (e.g. row1 = A1, B1, A2, B3, ..., A12, B12; row2 = C1, D1, C2, D2, ..., C12, D12). This is the same way the lrp plates were handled. This round of chip is labeled **ChIPrndC** on all of the eppendorf tubes containing samples

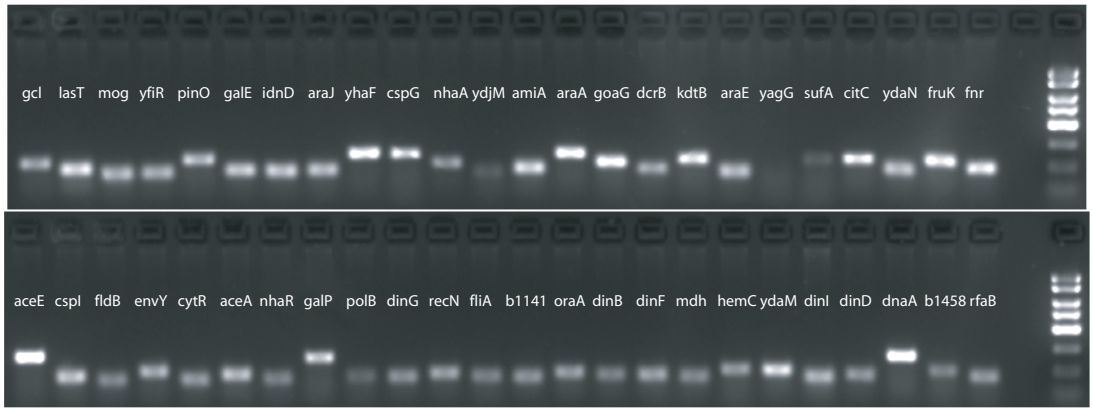


Figure 1.32: 300 ml, 1.5% agarose gel with 1.8 ul of 1% ethidium bromide run for 40 min at 100 volts. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used. These are the lexA targets. 8μl of the PCR/dye was loaded per lane.

and on the plates containing oligo mixes.

First round of CHIPrndC A302JWNU Tue Dec 20 17:48:04 EST 2005

fecIf (1) first then pdhRb (2)

No problems that I saw.

Second round of CHIPrndC A302JWNV Tue Dec 20 19:50 EST 2005

fecIe (3) first then pdhRd (4)

No problems that I saw.

Third round of CHIPrndC A302JWNW Wed Dec 21 17:49:49 EST 2005

pdhRa (5) first then fecIa (6)

No problems that I saw. I switched from the Marsh reservoirs to corning costar 4871 50ml reservoirs, and it is even easier to pipette the entire plate without running out of mastermix. I could probably use less extra master mix (but I'm not going to, better safe than sorry).

Fourth round of CHIPrndC A302JWNX Wed Dec 21 19:51:46 EST 2005

pdhRc (7) first then lexAd (8)

Fifth round of CHIPrndC A302JWNY Wed Dec 21 22:13:58 EST 2005

pdhRf (9) first then lexAc (10)

Sixth round of CHIPrndC A302JWO1 Thu Dec 22 00:15:48 EST 2005

lexA (11) first then lexA (12)

1536 qPCR reactions in one day is too much. I don't think I'll do this again anytime soon. Maybe I'll be puking SYBR green for xmas...

Seventh, Eighth, Ninth rounds of CHIPrndC Thu Dec 22, 2005 performed at 4PM, 6PM, 8PM.

A302JWO0 lexA (13) first then fecI (14)

A302JWNZ fecI (15) first then fecI (16)

A302JWNI pdhR (17) first then lexA (18)

Round Eight was run on the other ABI 7900HT machine.

It is finished Thu Dec 22 22:50:59 EST 2005

Brief Conclusions: Tue Jan 17 12:08:56 EST 2006

During the last round, lubing up the pipettor with silicon grease really got annoying, because there is a very fine balance between too much and too little grease. If the pipettor is not in that sweet spot, it is *very* frustrating to pipette. Getting uniform volumes requires throwing out reagent (I always find the sweet spot when I'm pipetting primers, since I have a very large supply of them) and adding or removing silicon until you hit the magic spot where the tips are all accurate, even and pipetting is sooooo much faster. It is stupid to have to deal with this unnessary problem. The problem to a large extent seems to be dependent on how warped the tip box becomes after autoclaving.

For the results above I used almost entirely Fisher tips (21-277-2B). They work well for standard pipetting in the lab with a Rainin P2 or P10. The last round I used boxes of finntip tips from Thermo Electron (the company making the multichannel). It was a dream come true, silicon woes were over and they stuck on the pipettor channels like flies on stink. No problems, always accurate, a little expensive. However, I did a little research today. By using the manufactors refill packs I can save the environment (a little) and the cost is almost the same as the fisher tips (3 one-hundredths of a penny more per tip; 0.0426 per tip (thermo) vs 0.0423 (fisher)).

One slight worry is the tip boxes containg 2x as many tips (192 instead of 96). Having 96 per box made it easy to keep track of location on the plate and not pipette into the wrong set of wells. 192 will speed things up, but make it more difficult to have a quick glance and immediately know where you are if someone stops by to talk to you. The are numbers are (9400327 for a refill and 9400326 for a starter kit that contains the 192 hole, tip boxes).

1.8 Sharing the Transcription Factor TOPO cds cloning primers

Thu Jul 13 14:20:01 EDT 2006

Hemali Patel (in our lab) needed to be able to amplify the full cds of many genes for some microarray spike-in studies. I gave here a 5 uM sample of 25 genes I thought I might use for the ChIP studies. I made and 200 μl plate for myself (in case I want to amplify these in the future) and gave 30 μl to here. The layout of the plate is below.

F indicates the forward primer; R indicates reverse primer.

Full cds primer plate #7136												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	nus F	gadW F	glcC F	flhC F	sfsA F	cbl F	pdhR F	nac F	yheO F	yrb F	fliA F	rhaR F
B	nus R	gadW R	glcC R	flhC R	sfsA R	cbl R	pdhR R	nac R	yheO R	yrb R	fliA R	rhaR R
C	gadX F	lrp F	hyaC F	lexA F	bolA F	yhiF F	abgR F	fecI F	ymfL F	sspA F	cytR F	yidP F
D	gadX R	lrp R	hyaC R	lexA R	bolA R	yhiF R	abgR R	fecI R	ymfL R	sspA R	cytR R	yidP R
E	ymfN F	-	-	-	-	-	-	-	-	-	-	-
F	ymfN R	-	-	-	-	-	-	-	-	-	-	-
G	-	-	-	-	-	-	-	-	-	-	-	-
H	-	-	-	-	-	-	-	-	-	-	-	-

Chapter 2

Towards a faster, more reliable ChIP protocol

THIS CHAPTER/PROJECT IS NEAR COMPLETION

Check out my blog post "Factorial and response surface optimization of a chromatin immunoprecipitation protocol" for a more in-depth introduction of my goals for this project. Also, if you have questions or comments, please post them on the blog as well.

In Chapter 1, I got a ChIP protocol hammered out well-enough to find a known targets to the transcription factors Lrp, PdhR, and FecI. The protocol is dreadfully slow and tedious however. If I could shorten it and produce at least as good separation between enriched and random then adding more replicates would be less painful and would produce more reliable results.

The hope is that by doing a couple rounds of factorial experiments with 4-8 factors, I'll be able to provide *better* enrichment in *much less* time. Hopefully, filter based bead-washing methods or dynal magnetic beads will pan-out, allowing me to remove one of the easiest places to make an error (i.e. sucking your beads out of the tube). After getting the protocol shortened, might be possible to use response surface methods to increase enrichment even further? Would be a cool tech paper if it worked.

The methods I'm using to shorten/improve this protocol are described in the excellent book: *Statistics for Experimenters. Box, Hunter, Hunter.* If you like history and old books, Fisher developed and wrote quite a bit about the experimental designs I use in this chapter.

Brief Update *Thu Dec 13 15:10:17 EST 2007*: Two much improved protocols have resulted from this chapter: one requiring 1.5 days and one requiring 2.5 days. Both of the protocols are 96-well format capable. The protocols can be found on J's Blog in the post: "Optimized ChIP Protocols" (http://blog-di-j.blogspot.com/2007/12/optimized-chip-protocols.html).

Brief Update *Wed Aug 23 18:17:53 EDT 2006*: I originally planned to use the more enriched DNA from this protocol for SAPE. Now I hope to be able make a Paired-end-tag library to just sequence it straight away with highly parallel sequencing.

Here are some things I'd like to achieve:

- Shrink to 6ml in 48-well plate; shrink other downstream volumes accordingly
- try filter method instead of pelleting to wash beads
- Move to filtrex 96-well
- optimize antibody concentration
- optimize formaldehyde conc
- optimize cross-linking time
- try other chemical that aids protein complex formation (David showed me this)
- move to diagencode 12-at-a-time shearing
- remove preclear
- switch to qiagen PCR cleanup to cleanup DNA
- incubate with antibody for shorter time
- incubate with beads for shorter time
- if time shortens work, qiagen cleanups work and filters work: it might be possible to get quantification part 1, quantification part 1, immunoprecipitation, immunoprecipitation, and immunoprecipitation bead washing into one day; taking the 4-day protocol down to three, with the last day being really short, short enough that the qPCR rxns could start that day.
- try using magnetic beads instead of agarose; would remove idea of using filtrex, instead would use a 96-well dynal magnetic plate

2.1 ChIP improvement round 1: shortening the time

This round will mainly focus on seeing if shortening a few bottleneck steps without worsening performance [1]. A few things that might improve the noise enrichment vs true enrichment ratio have been include as well and are 1) altering antibody conc and 2) altering the formaldehyde concentration

Here is the factor list (the 0-state is what I've used in the past):

[1] this is nice because normally when I do statistical tests I'm looking to improve something; now I just want to make sure things aren't getting any worse

factor	0-state	1-state
plasmid	low copy	high copy
formaldehyde conc.	1%	0.1%
quench with glycine	yes	no
quantification part 1 use qiagen (requires changing volumes)	no	yes
shearing	4 x 20% x 30 sec	3 x 10% x 20 sec
preclear	yes	no
antibody conc (per 25 μg DNA)	2 μg	6 μg
antibody incubate time	overnite	1hr
bead incubate time	2hr	30min
wash method	pellet	filter column
final cleanup	phenol:chloroform	qiagen
bead type	agarose	dynal

Other than these factors, all other aspects will remain as similar to the protocol in section C.3 as possible.

2.2 First round ChIP optimizations

I'm doing a 16 sample factorial design with lrp (the TF I have cloned with the most known targets). The factors are (low — high):

1. IPTG (0.01 μM — 1.0 μM)
2. formaldehyde (0.1% — 1%)
3. quench with glycine (yes — no)
4. shearing time (4 x 20% x 30 secs — 1 x 10% x 30 secs)
5. preclear with beads (yes — no)
6. antibody concentration per 25 ug of starting material (2 μg — 10 μg)
7. incubation time with beads prior to washing (10 min — 2 hr)
8. bead type (agarose — dynal)

Along with this, I'm also going to test using Charge Switch or Qiagen PCR cleanup rather than phenol/chloroform for the initial DNA quantification step. If they are comparable, the charge switch and Qiagen columns are certainly much faster (and safer). I'm not sure how good they are with small quantities, because perhaps it could be used in the final elution too? Maybe in round 2.

2.2.1 Phenol -vs- ChargeSwitch -vs- Qiagen PCR purification for cleanup of sheared, decrosslinked DNA

While I'm waiting for reagents to come in for the factorial experiment, I'm going to test to see if it matters which cleanup kit I use in the first step where I quantify my amount of starting DNA before adding the antibody.

Wed Apr 25, 2007

I spread some lrpB-TOPO-antiExpress (lrpB) onto an amp/agar plate.

Sun Apr 29, 2007

I picked two colonies (lrpB:J1 and lrpB:J2) to grow overnight in 4 ml of LB+amp.

Mon Apr 30, 2007

I grew a 1:100 dilution of lrpB:J1 and lrpB:J2 from the overnight cultures in 25 ml of LB + amp. After 1 hr, I added 1:1000 of 1M IPTG (1 μM).

As the cells reached an OD600 of *, I then followed the ChIP protocol for crosslinking, lysing, and shearing the chromatin. Minor modifications were: the initial cell pellet centrifigation I ran for 15 min rather than 10 min (cells then precipitated completely), after the addition of 500 μl of 2x Pallson IP buffer I incubated 5 min and only spun at 100 rpm (lysis still went to completion with slower rpm and shorter time).

For the shearing, I sheared J1 3 x 20% x 30 seconds, and I sheared J2 3 x 20% x 30 seconds.

To reverse the crosslinks, I placed 10 μl , 50 μl , and 100 μl of sheared DNA into a total volume of 125 μl (topped up with 115 μl , 75 μl , and 25 μl of H_2O respectively) with 0.5 μl , 2.5 μl , and 5 μl of Proteinase K [Ambion] respectively. I did this for both J1 and J2. I made three replicates of each of these, so I'd have one replicate for each of: ChargeSwitch, Phenol, Qiagen PCR cleanup column. Total number of tubes was 2 (J1, J2) x 3 (10 μl , 50 μl , 100 μl) x 3 (ChargeSwitch, Phenol, Qiagen) = 18 tubes. I did the dilutions to try and check the linearity of each method as the concentration of DNA changed (I'm actually concerned more with precision than accuracy, so I'd rather the slope be accurate than the intercept).

Tue May 1 19:55:55 EDT 2007

I ran 6 cleanups for all of the cleanup kits according to the manufacturer's instructions, except that for the ChargeSwitch, they recommend a 25-50 μl PCR reaction and my starting volume (besides not being from a PCR) was 125 μl . I adjusted the concentration of the initial binding purification reagent accordingly (they wanted a 1:1 ratio). I eluted each reaction into 30 μl . The yields were:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**
J1 : 10 p	21.2	1.98	1.74
J1 : 10 c	29.3	2.01	1.96
J1 : 10 Q	28.3	1.75	1.25
J2 : 10 p	28.9	2.06	2.10
J2 : 10 c	32.6	1.88	1.91
J2 : 10 Q	35.9	2.50	1.28
J1 : 50 p	123.4	1.98	2.05
J1 : 50 c	64.7	1.99	2.00
J1 : 50 Q	121.2	1.85	1.25
J2 : 50 p	79.0	2.10	2.27
J2 : 50 c	75.2	1.99	2.00
J2 : 50 Q	149.9	1.77	1.38
J1 : 100 p	663.9	1.95	2.03
J1 : 100 c	39.8	1.93	1.54
J1 : 100 Q	275.6	1.81	1.67
J2 : 100 p	1112.2	2.09	2.27
J2 : 100 c	54.8	1.99	1.81
J2 : 100 Q	299.6	1.81	1.62

(J1 = lrpB:J1 = sheared 3x; J2 = lrpB:J2 = sheared 1x; 10, 50, 100 = starting dilution of sheared DNA; (p, c, Q) = (phenol:chloroform, ChargeSwitch, Qiagen PCR cleanup column).

It's a lot easier to see what's going on with those numbers when we plot them in Figure 2.1.

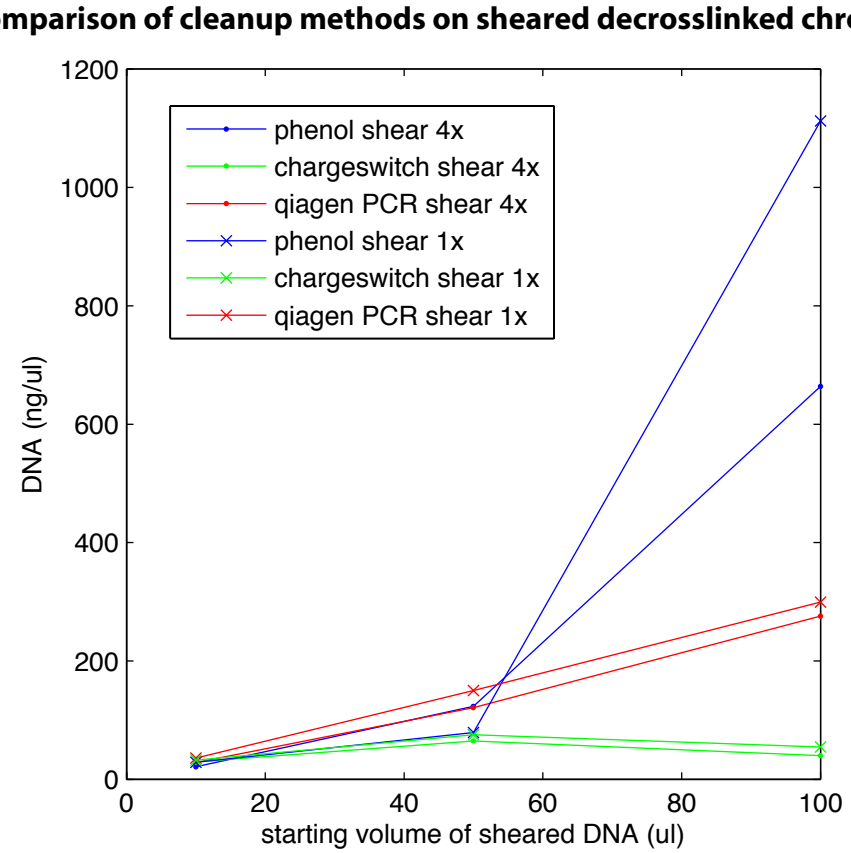


Figure 2.1: The qiagen PCR cleanup kit is the only DNA cleanup method that was linear across the dilution range

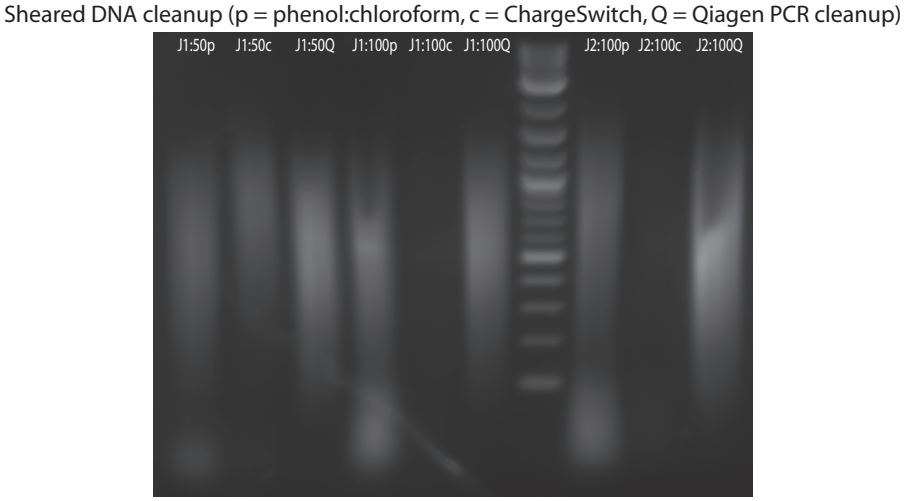


Figure 2.2: The phenol:chloroform extraction lanes retained the tiny DNA fragments in the high volume purification, which may have led to the strange non-linear scaling of the measured concentrations.

Brief Conclusions: The Qiagen concentration estimates are spot-on (Figure 2.1), but DNA purity from this kit is the worst (see table above). Not only is the Qiagen yield linear with the dilution amount, the slope is also about correct (i.e. when dilution is 1/2 the conc is 1/2). Looks like big phenol numbers are from the little pieces that the other kits remove (Figure 2.2). Also notice that although chargeswitch seems to blow, it does do a really nice job of cutting off the size of the DNA at a larger size than the other kits. I should keep this in the back of my mind, because this could be a really useful way to avoid gel purification to remove adaptors and stuff like that. They claim in the manual that the size cutoff is adjustable, so it I can push the size high enough, I might not even need to gel select the cDNA step. Last, I ordered a Qubit from Invitrogen. This is a little machine that makes using their DNA/RNA quantification kits easy (e.g. picogreen). I want to try this dilution test again, but down to much lower levels to see about the possibility of using Qiagen rather than EtOH for the final DNA cleanup after the immunoprecipitation.

In summary, it looks like the Qiagen kit is the way to go for the initial quantification of sheared DNA yield. Based on some results from Henry Lee in the Collins lab, in the next round where I try diluting further, I might quantify all of the samples with picogreen to prevent the lessen influence of mRNA in the spec readings.

2.2.2 first round ChIP optimizations: detailing the plan

Mon May 7, 2007

The factors and reasoning behind the factors was detailed above at the start of this section (i.e. section ChIP improvement round 1). Below is the table that I'm actually using to experimentally pursue this goal. The rows were randomized with the rand_perm() function in matlab.

As far as nomenclature goes, I'll be referring to the samples by their randomized order. For each sample N there is NA, NB, and NC where A = positive control sheared DNA, B = no antibody negative control, and C = antibody enriched (hopefully) sample.

Randomized 8-factor ChIP optimization 16-sample fractional factorial design								
randomized order	**IPTG**	**form**	**quench**	**shear**	**pre**	**anti**	**incubate**	**bead**
1	0.01 μM	0.10%	yes	4x20%x30	yes	10 μg	10min	agarose
2	0.01 μM	1%	yes	4x20%x30	no	2 μg	10min	dynal
3	0.01 μM	0.10%	yes	1x10%x30	yes	2 μg	2hr	dynal
4	0.01 μM	1%	no	4x20%x30	yes	2 μg	2hr	agarose
5	1 μM	0.10%	no	4x20%x30	yes	2 μg	10min	dynal
6	1 μM	0.10%	no	1x10%x30	yes	10 μg	2hr	agarose
7	0.01 μM	0.10%	no	1x10%x30	no	2 μg	10min	agarose
8	0.01 μM	1%	yes	1x10%x30	no	10 μg	2hr	agarose
9	1 μM	1%	yes	4x20%x30	yes	10 μg	2hr	dynal
10	1 μM	0.10%	yes	1x10%x30	no	10 μg	10min	dynal
11	0.01 μM	0.10%	no	4x20%x30	no	10 μg	2hr	dynal
12	0.01 μM	1%	no	1x10%x30	yes	10 μg	10min	dynal
13	1 μM	0.10%	yes	4x20%x30	no	2 μg	2hr	agarose
14	1 μM	1%	no	1x10%x30	no	2 μg	2hr	dynal
15	1 μM	1%	yes	1x10%x30	yes	2 μg	10min	agarose
16	1 μM	1%	no	4x20%x30	no	10 μg	10min	agarose

2.2.3 first round ChIP optimizations: growing, shearing, lysing, sonicating

Mon May 7, 2007

I performed this step similarly to the one I did for J1 and J2 in section 2.2.1. This was the first time I had used the glycine to quench the crosslinking. I grew the cells for 3 hrs and 20 minutes before taking the 15 ml samples for crosslinking (background subtracted OD600 is in the table below). I used 750 μl of 2.5 M glycine (1/20) in the 15 ml reaction. Based off the information from that section, I cleaned up the reactions with Qiagen, and I obtained the following yields:

Sheared DNA yields from ChIP factorial optimization round 1							
Sample ID	**ng/uL**	**260/280**	**260/230**	**num μl for 25 μg**	**μl buffer for 1:10 dilution**	**total volume**	**OD600**
ChIP 1	242.58	1.77	1.37	103	928	1031	0.406
ChIP 2	300.41	1.74	1.28	83	749	832	0.438
ChIP 3	289.23	1.97	1.63	86	778	864	0.401
ChIP 4	276.1	1.74	1.45	91	815	905	0.417
ChIP 5	300.52	1.71	1.09	83	749	832	0.368
ChIP 6	275.34	1.94	1.63	91	817	908	0.365
ChIP 7	283.41	1.99	1.55	88	794	882	0.408
ChIP 8	268.74	1.8	1.46	93	837	930	0.398
ChIP 9	268.68	1.8	1.48	93	837	930	0.367
ChIP 10	275.22	1.93	1.64	91	818	908	0.366
ChIP 11	279.2	1.82	1.62	90	806	895	0.402
ChIP 12	219.63	1.77	1.01	114	1024	1138	0.397
ChIP 13	248.9	1.8	1.51	100	904	1004	0.389
ChIP 14	257.39	1.89	1.59	97	874	971	0.377
ChIP 15	219.73	1.96	1.47	114	1024	1138	0.370
ChIP 16	305.7	1.77	1.33	82	736	818	0.344

To Do!!! I still need to run these on a gel to check the shearing range. I also should check my previous experiments to determine how much to run on the gel

May 15, 2007

I ran the 16 factorial samples on two 1.5% agarose gels (Figure 2.3).

Brief Conclusions: Well it's too late now (I've already run the qPCR samples), but if I had this shearing to do this factorial again, I'd make the shearing for the low-shear a little longer or stronger. The low-shear lanes have the DNA average at around 1kb or more. The high-shear lanes have the DNA average length at around 350bp.

2.2.4 first round ChIP optimizations: antibody, beads, washing, crosslink reversal

Tues May 8, 2007

Notes1: sample 3 took forever for beads to finish binding to the magnet; sample 6, 7, 10 were very viscous (I didn't notice that well with sample 7 and for that one a lot of the beads were sucked out of the tube)

Notes2: in samples 7B and 7C the beads are almost gone; 10C is a complete impenetrable ball of magnetic beads. The washes and 5 minute rotations have little effect on breaking up this ball

Notes3: I prepared the dynal beads using a similar strategy to the Young Lab protocol in Nature Protocols. 1) add N μl of beads; 2) collect with magnet; 3) add 15 x N μl of block solution (0.5% BSA in PBS); 4) repeat steps 2 and 3; 5) resuspend beads in N μl of block solution

Shearing results from lrp fractional factorial round 1

ladder is the NEB 2-log ladder

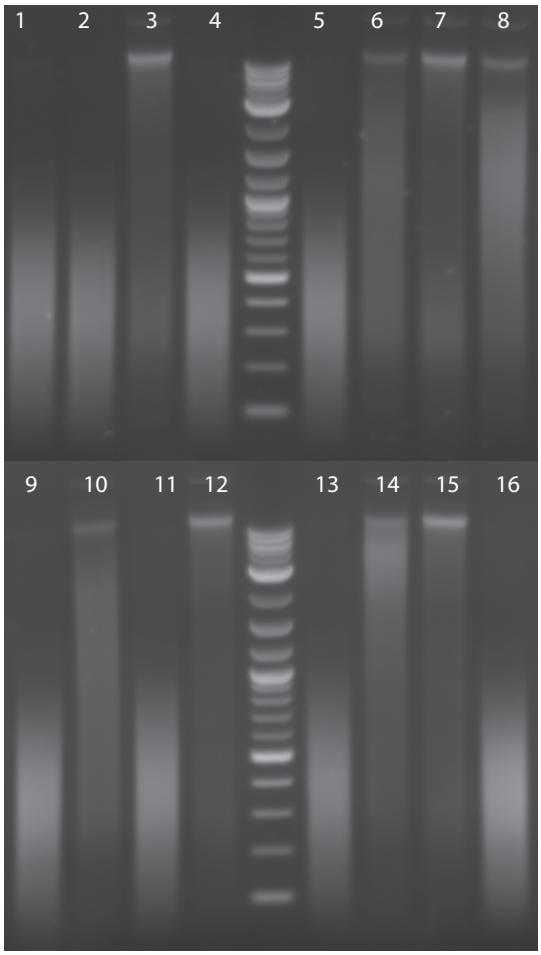


Figure 2.3: 1.5% gel; 16 sheared samples; the lightly sheared samples were 1x10%x30secs. the others were 4x20%x30secs.

I used the same volume of beads for dynal and agarose: 40 μl preclear and 60 μl immunoprecipitation. The washings were done in the cold room. They were done according to my original chip protocol except some of the samples were incubated with antibody and beads 2 hrs each (4 hours total) and other samples were incubated 10 min each (20 minutes total).

For the elution, I used the Young protocol (which used something similar to TE + 1% SDS) for the dynal elution. And I used the original protocol for the agarose elution. Crosslinks were reversed overnight in a water bath at 65C.

2.2.5 first round ChIP optimizations: purification of DNA products

Tues May 9, 2007

I added 1 μl of proteinase K to all 32 samples. To the dynal samples I added H_2O to make them be at the same volume as the agarose samples (450 μl total). I added Tris and EDTA to the agarose samples, as per my original ChIP protocol.

I did a phenol extraction of all of the samples using Gel Phase lock (light) tubes. I did not do a second chloroform extraction. I added 1 μl of glycoblue to the EtOH precipitations. I did the

EtOH precipitations in three rounds hoping that would help prevent having the DNA pellets come unstuck from the tubes. I don't think it mattered. Next time I'd just do two rounds of 16.

2.2.6 first round ChIP optimizations: preparing and testing the ChIP primers

I'm using the same 12 random primers from last time, even though I think know at least one of those primer doesn't work. 11 genes is still a fairly large negative sample for comparison with my enriched genes. The row order was randomized, as was the gene order within each row. The twelve genes chosen for lrp include:

- primers to test the effect of shearing range on precision (serA5KB and serA1KB, where are 5KB and 1KB upstream respectively from the known lrp binding site in front of serA).
- known targets that were enriched in my samples for the PLoS 2007 CLR paper (gtlB, serC, leuL, serA, livK)
- a known target that wasn't enriched in my samples for the PLoS 2007 CLR paper (stpA); perhaps this is a false positive in RegulonDB
- new CLR identified lrp targets and enriched from the PLoS paper (yhjE and pntA); perhaps these were false negatives in RegulonDB
- CLR identified lrp targets that were NOT enriched from the PLoS paper (aroP and metA); perhaps these were false negatives in CLR

Lrp factorial primer plate												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	gcl	mog	pinO	idnD	yhaF	nhaA	aimA	goaG	kdtB	yagG	citC	fruK
B	yhjE	gtlB	stpA	serC	leuL	serA	aroP	pntA	metA	serA5KB	serA1KB	livK

I made one 300 μl plate and one 100 μl plate. Both plates had primers at 2 μM concentration.

To make sure the primers still work (they're about 1yr old) and that I set up the primer plate correctly, I did a 20 μl PCR with the cheap NEB Taq and genomic DNA. I ran them out on a 2% agarose gel (Figure 2.4).

Brief Conclusions: The random primer gel is a little wacko, but everything looks pretty good. mog (random lane2) and yagG (random lane 10) were wacky in the previous ChIP study too, so I'm not worried about those guys.

testing the primers with qPCR master mixes

Sat May 12, 2007

The DynamoHS SYBR green master mix from finnizymes is about 1/3 cheaper than the ABI one, so I'm going to test that one and the ABI to compare the results.

Raw data in excel format.

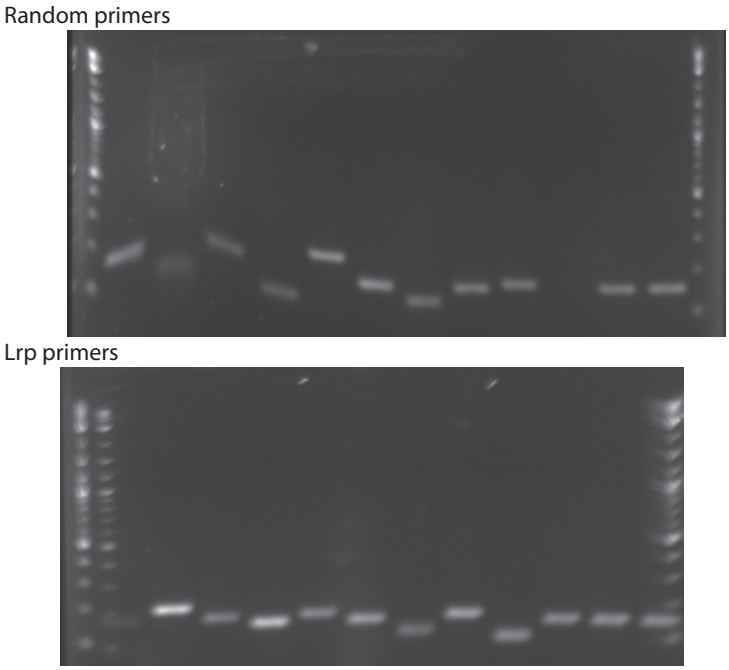


Figure 2.4: 2% gel of the random primers and lrp primers to be used in the factorial ChIP optimization experiment. The identity of the lanes is the same as in the table Lrp factorial primer plate table above.

Brief Conclusions: The results were pretty similar besides a few outliers. One thing that you cannot see (but that I saw in the ABI software) is that the primer dimer formation as indicated by the final DNA melting curve was less with Dynamo. The dynamo also produced a little less signal over all, and it had a mildly worrying feature that after the signal saturation point was reached, the signal actually began decaying a little (the ABI master mix rises slightly after PCR saturation). The Ct values, which are in the area of the curve that really matters, were fine though.

2.2.7 first round ChIP optimizations: dynamoHS qPCR

Sun May 13, 2007

I ran samples 1-8 B and C (16 total) with all 24 genes (384 samples total) using 10 μl dynamo HS master mix, 5.5 μl H_2O , 0.4 μl ROX (1x final concentration), 1.5 μl primer (2 μM stock), and 3 μl template (as in the previous experiments for the PLoS paper, I diluted the 100 TE+template with 100 μl of H_2O for a final volume of 200 μl). I made enough master mix for 10 extra rxns and ran short 19 rxns (that brings back bad memories!). I ordered a matrix multichannel electronic pipettor that should help alleviate this problem in the future. The 19 rnxs I ran short of master mix on, I made a second batch of master mix for and filled them by hand. The plate barcode was A302JWNB. As far as the 384-well plate organization goes. A1 (and all odd columns) of the qPCR plate contains A1...A12 of the lrp plate (the random row), A2 (and all even columns) of the qPCR plate contains B1...B12 of the lrp plate. For the template samples 1B...8B was put in rows P,N,L...B; samples 1C...8C was put in rows O,M,K...A.

Tues May 15, 2007 I ran samples 9-16 B and C (16 total). The reaction concentrations were the same as above. I made enough master mix for 30 extra rxns and did NOT run short this time. The

Ct values from ABI and Dynamo HS			
-	Ct ABI	Ct Dynamo	Difference
gcl	14.995	15.037	0.042
yhjE	14.542	14.902	0.360
mog	20.229	15.671	-4.558
gtlB	15.562	15.320	-0.242
pinO	19.173	16.111	-3.062
stpA	14.838	15.108	0.271
idnD	15.320	17.030	1.709
serC	14.234	14.401	0.167
yhaF	18.799	18.801	0.002
leuL	15.099	15.330	0.231
nhaA	14.921	15.169	0.248
serA	14.430	14.430	0.001
aimA	15.426	16.114	0.688
aroP	14.129	14.819	0.690
goaG	16.053	16.179	0.126
pntA	14.987	14.810	-0.177
kdtB	16.370	15.821	-0.549
metA	13.999	14.327	0.327
yagG	Undetermined	36.849	Undefined
serA5KB	14.098	14.511	0.413
citC	17.118	17.105	-0.013
serA1KB	14.607	14.612	0.004
fruK	17.447	17.744	0.297
livK	14.654	14.721	0.067

plate barcode was A302JWNC.

Also, I did a quick skim of the qPCR data for the first eight experiments. I identified by eye the three samples that I felt were noticeably better than all of the other experiments: all had 1% formaldehyde in common (rather than the 0.1% formaldehyde).

Here is the raw data for the hsDynamo qPCR rxns factorial rnd1 samples 1-8 and factorial rnd 1 samples 9-16.

2.2.8 first round ChIP optimizations: ABI master mix qPCR

I'm trying the exact same qPCR reactions as above but with the ABI master mix that we pay about a third more for and that I used in the PLoS netinfer paper ChIP experiments. Since ROX is included in the master mix I added water in place of the 1:50 dilution of ROX in the above dynamoHS experiments.

Wed May 16, 2007

I ran the first qPCR plate with the ABI mix.

Tues May 22, 2007

I ran the second qPCR plate with the ABI mix.

Here is the raw data for the ABI qPCR rxns factorial rnd1 ABI samples 1-8 and factorial rnd 1 ABI samples 9-16.

2.2.9 Summary of first round results

Summary results from lrp factorial first round						
		hsDynamo qPCR		ABI qPCR		
id	*factor*	*effect*	*pval (Lenth)*	*effect*	*pval (Lenth)*	*conclusion*
1	IPTG	-0.1787	0.031	-0.3665	0.012	high IPTG may be slightly worse
2	formaldehyde	1.0555	0.0002	1.1803	0.0005	form. concentration is very important (1% >> 0.1%)
3	quench	0.6774	0.0007	0.5972	0.0021	quenching definitely helps
4	shear	0.0271	0.6357	-0.1646	0.13	shearing alone is not significant
5		-0.0362	0.7402	0.1825	0.1633	
6		-0.2359	0.0119	0.1192	0.2048	
7		-0.0492	0.9181	0.0393	0.6178	
8	form/quench?	-0.7273	0.0005	-0.8768	0.0011	if you have low formaldehyde - don't quench
9	form/shear?	1.0807	0.0002	1.1853	0.0005	if you have high formaldehyde - shear more
10		0.4694	0.0016	0.5356	0.0045	
11	preclear	-0.053	0.9784	-0.0151	0.4762	
12	antibody conc	-0.0532	0.9819	0.032	0.5709	
13	incubation time	0.3245	0.0053	0.4424	0.0098	2hr incubation is a little better than 10 minutes
14	bead type	-0.65	0.0007	-0.9791	0.0008	dynal beads aren't as good as agarose
15		-	1	-	1	

Brief Conclusions: Mon Jun 18 14:25:42 EDT 2007

I'm just going to summarize my gut feelings based on the above table.

Main effects (linear). First, lower protein concentration might be important. Perhaps, having lrp on a high-copy plasmid with high concentration of IPTG for induction just overwhelms the genome and the lrp binds everywhere. Second, formaldehyde concentration is the single most important factor screened in this round. The 1% was much more enriching than the 0.1%. Presumably, the 0.1% is just not enough to bind everything together in those 10 minutes. When I picked the best enrichments by eye before performing the computational analyses of this data the three best I choose all had 1% formaldehyde. Third, quenching with glycine helps. Halting the crosslinking should improve the consistency and doesn't add much extra work. Fourth, shearing did not significantly effect the results. Fifth, preclearing didn't make a difference. Sixth, adding a lot of extra antibody didn't help. Seventh, 2x2hr incubations was a fair amount better than 2x10 minute incubations. Eigth, dynal was not quite as good as the agarose beads.

Combinatorial effects (nonlinear). Unfortunately the combinatorial effects are all confounded with each other. I'm going to make my best guess at the 2-factor interactions that could explain the combinatorial interactions we see in the table above. My main guess is that since formaldehyde was the most important main effect, the combinatorial effects were probably in some way related to formaldehyde. So I'm guessing the non-linear effect number 8 is an interaction between formaldehyde/quenching. If that's the case, the effect would be interpreted as lack of quenching is beneficial when you are using 0.1% formaldehyde. This hypothesis seems pretty reasonable, because 0.1% didn't work as well as 1%, so the longer you crosslink with the low concentration, the closer you are to the shorter incubation at the high formaldehyde concentration. The second non-linear interaction I'm guessing is an interaction between shearing and formaldehyde. If so, it seems to be beneficial to shear more for higher concentrations of formaldehyde, presumably because the higher formaldehyde concentrations bind everything up together.

What I'm a gonna do.

1. use low IPTG (I'm also going to clone the tagged lrp into a low copy plasmid to bring the expression down even further)
2. quench - this I'll do from here on. it takes a trivial amount of additional work. and seems to help things out
3. keep formaldehyde concentration high. I want to explore this further in the next round with higher formaldehyde concentrations
4. shear fairly well. seems to help with higher formaldehyde. if not, it doesn't hurt (though it does take extra time). I also want to explore the interaction with formaldehyde later in a response surface experiment
5. don't preclear - needless waste of time
6. antibody concentration - higher didn't matter; I'm going to try lower in the next round
7. use faster incubation time - I know, 2 hr was significantly better, but it wasn't drastic. Maybe I should explore this in more detail to see what really matters.
8. use dynal beads - again, I know this is in contradiction to the results, but it is *much* faster to use the dynal beads, and it's easier to be consistent, so I feel I'll have less noise if I go with the dynal
9. use hsDynamo master mix from Finnizymes (distributed by NEB); about half the cost and the results are very similar

2.3 scratchNotes

Ideas for next optimization round: 1) mix beads and antibody together during the DNA clean up/quanitification; 2) use Qiagen cleanup at the end (rather than EtOH); clone into ilaria's vector with the very low copy number

Based on factorial, go with sample 15 (randomized sample 2) as the default: 0.01uM IPTG, 1% formaldehye, shear 4x20%x30, no preclear, 2 μg antibody, incubate 10 minutes, dynal beads

In first round, formaldehyde was the biggest factor. Quenching was helpful. And there was a nonlinear effect that you needed to shear more with increased formaldehyde concentration.

spec with qubit?

Fo the second round,

2.4 Second round ChIP optimizations

summarize first round here. justify second round here.

2.4.1 second round ChIP optimizations: detailing the plan

2nd round the factors are (low — high):

1. starting DNA conc. (20 μg — 80 μg of DNA); need to grow in 50 ml flask
2. formaldehyde (0.5% — 4%)
3. lyse (normal — lyse in dilution buffer via sonication)
4. shearing time (6 x 20% x 30 secs — 2 x 20% x 30 secs)
5. antibody (0.25 μl — 2 μl)
6. bead concentration (10 μl dynabeads — 100 μl dynabeads)
7. wash (normal 3 salt wash + 2xTE — 2x TE only)
8. final DNA cleanup (phenol — qiagen PCR cleanup)

For the high DNA concentration, I plan to grow 60 ml of cells and lyse 50 ml (rather than my typical 25 ml grow 15 ml lyse).

For the lyse in dilution buffer, I plan to lyse the cells in dilution buffer during sonication (i.e. let the sonication lyse the cells rather than using lysozyme and high conc. sucrose) rather than the normal chemical lysis.

2.4.2 second round ChIP optimizations: checking the sonication

Mon May 21, 2007

I was a little surprised to see how not-sheared my DNA was for the low shear factor in round 1 of the optimization (Figure 2.3). This round I'm going to check the shearing range a little in a pretest just to make sure the shearing is a little more reasonable (even though the low shear seemed to work).

Specifically, I wanted to see the effect of using no lysis and doing the sonication in IP buffer. I made two sample culture volumes. Consistant with the volumes necessary or the factorial factor of 20 μl vs 80 μg of starting DNA, I grew 15 ml and 60 ml respectively. I also tested two formaldhyde concentrations (0.5% vs 4%) and two shearing lengths (2x20%x30 secs and 6x20%x30 secs), as it was found that there was a nonlinear interaction between these in the first round, and because I'm not sure how these DNA concentrations and formaldehyde concentrations will influence the shearing size.

The six 50 ml samples were taken from individual flasks with 70 ml starting volume of culture. The six 15 ml samples were taken from one 70 ml sample and the leftover culture from the 50 ml experiments (20 ml was left in the 50 ml cultures). The cells were grown for 3hr 30 min to an OD600 of around 0.45.

I ran the following 12 samples:

sample	sample vol.	formaldehyde conc.	chemical lysis	sonication strategy
1	15 ml	0.5%	no	2x20%x30sec
2	15 ml	4%	no	2x20%x30sec
3	15 ml	0.5%	no	6x20%x30sec
4	15 ml	4%	no	6x20%x30sec
5	50 ml	0.5%	no	2x20%x30sec
6	50 ml	4%	no	2x20%x30sec
7	15 ml	0.5%	yes	2x20%x30sec
8	15 ml	4%	yes	2x20%x30sec
9	50 ml	0.5%	yes	2x20%x30sec
10	50 ml	4%	yes	2x20%x30sec
11	50 ml	0.5%	yes	6x20%x30sec
12	50 ml	4%	yes	6x20%x30sec

Tues May 22, 2007

The 12 samples were cleaned up with a Qiagen PCR cleanup kit and eluted into 30 μl of EB buffer. The yields from the nanodrop were:

Sheared DNA yields from shearing test of factorial round 2			
Sample ID	**ng/ul**	**260/280**	**260/230**
1	228.11	1.8	1.58
2	171.32	1.77	1.52
3	234.09	1.71	1.11
4	183.86	1.81	1.57
5	474.14	1.81	1.8
6	487.99	1.78	1.64
7	206.36	1.82	1.66
8	165.32	1.74	1.16
9	420.99	1.82	1.87
10	351.9	1.81	1.75
11	429.26	1.81	1.89
12	458.31	1.8	1.86

3 μl of sheared DNA was run on a 1.5% agarose gel (Figure 2.5).

Wed May 23, 2007

I also spun down all 12 samples at max speed for 3 minutes to see if there was any more cellular junk for the sonicated only -vs- the chemical lysis + sonicated samples. None of the 12 samples had a noticable precipitate.

Brief Conclusions: I like these results better than what I saw above for the round1 shearing checks (see Figure 2.5 vs the round one shearing in Figure 2.3). There isn't a huge difference between these samples, but the 6x sheared are noticably shorter. Note that the quantities of DNA in each lane aren't the same for the 50 ml samples and the 15 ml because the yields were different and I just ran 3 μl for all samples. By eye, I don't see any noticable difference between the samples that were chemically lysed and those that were lysed as part of the sonication process.

2.4.3 second round ChIP optimizations: factors

I'm pretty much using the protocol I wrote up from the last round (see section C.4 on page 417). I'm just adding in the factors above.

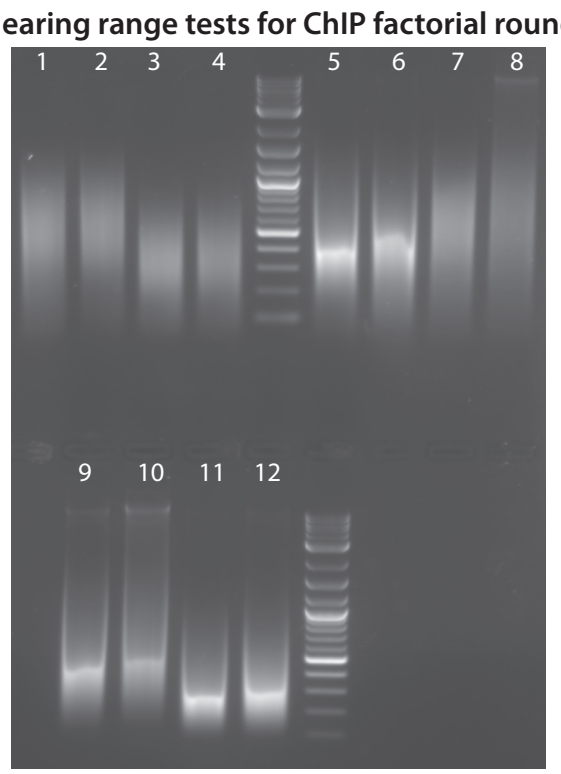


Figure 2.5: shearing range tests with different concentrations of DNA and formaldehyde, different shearing amounts, and different buffers. gel is a 1.5% agarose gel.

randomized setup for factorial round2									
randomized order	**DNAconc**	**form**	**lyse**	**shear**	**anti**	**beadConc**	**wash**	**DNAclean**	**factorialOrder**
1	20ug	4%	no	2x20%x30	2ug	100ul	normal	qiagen	3
2	80ug	0.50%	yes	6x20%x30	0.25ug	10ul	2xTE only	phenol	14
3	20ug	0.50%	no	6x20%x30	0.25ug	100ul	2xTE only	qiagen	9
4	80ug	4%	yes	6x20%x30	2ug	100ul	2xTE only	qiagen	16
5	80ug	4%	yes	2x20%x30	2ug	10ul	normal	phenol	8
6	80ug	4%	no	6x20%x30	0.25ug	100ul	normal	phenol	12
7	80ug	4%	no	2x20%x30	0.25ug	10ul	2xTE only	qiagen	4
8	20ug	4%	yes	2x20%x30	0.25ug	100ul	2xTE only	phenol	7
9	80ug	0.50%	no	2x20%x30	2ug	100ul	2xTE only	phenol	2
10	80ug	0.50%	no	6x20%x30	2ug	10ul	normal	qiagen	10
11	20ug	4%	yes	6x20%x30	0.25ug	10ul	normal	qiagen	15
12	80ug	0.50%	yes	2x20%x30	0.25ug	100ul	normal	qiagen	6
13	20ug	0.50%	yes	2x20%x30	2ug	10ul	2xTE only	qiagen	5
14	20ug	4%	no	6x20%x30	2ug	10ul	2xTE only	phenol	11
15	20ug	0.50%	no	2x20%x30	0.25ug	10ul	normal	phenol	1
16	20ug	0.50%	yes	6x20%x30	2ug	100ul	normal	phenol	13

The randomized matrix of factors is available here: in excel format.

2.4.4 second round ChIP optimizations: growing, shearing, lysing, sonicating

Sun May 27, 2007

I grew the cells up as before (adding IPTG 1hr after the 1:100 incubation).

OD600 (background subtracted) and sheared DNA yields were:

OD and sheared DNA yields for lrp ChIP factorial round2					
Sample ID	**OD600**	**ng/ul**	**260/280**	**260/230**	**10ug or 25 ug**
1	0.474	152.41	1.89	1.74	131
2	0.467	348.89	1.83	1.67	143
3	0.462	172.9	1.92	1.92	116
4	0.468	373.37	1.84	1.66	134
5	0.465	383.56	1.81	1.59	130
6	0.46	378.22	1.83	1.81	132
7	0.456	334.07	1.94	1.8	150
8	0.477	183.45	1.89	1.8	109
9	0.494	372.1	1.85	1.66	134
10	0.482	480.82	1.43	0.67	104
11	0.507	253.12	1.7	0.9	79
12	0.501	335.63	1.84	1.68	149
13	0.511	313.93	1.7	0.89	64
14	0.491	181.65	1.89	1.91	110
15	0.493	184.1	1.91	1.93	109
16	0.486	175.03	1.88	1.78	114

Notice in the table above I include the amount needed to get 10 μg and 25 μg . In the original plan, I was going for 20 μg and 80 μg , but I realized that I forgot to consider that I was eluting into 1/2 of the volume after I cleaned up the sheared DNA, so in the past I've been using 10 μg NOT 20 μg of DNA (opps). It was impossible to get up to 80 μg because that wouldn't fit into my tube, so I went with the biggest number that would fit, which was 25 μg .

Note this time instead of using slightly different amounts of sheared DNA for immunoprecipitation, I just used the median volume for that concentration of DNA (i.e. 10 μg or 25 μg). The median volume was 134 μl sheared-crosslinked DNA and 1207 μl Dilution Buffer for the 25 μg samples and 110 μl sheared-crosslinked DNA with 986 μl Dilution Buffer for the 10 μg samples.

2.4.5 second round ChIP optimizations: antibody, beads, washing, crosslink reversal

Mon May 28, 2007

everything went fine here just a simple note: 6C and 6B were slower to bind the magnet (presumably they were more viscious) than 1C or 1B.

2.4.6 second round ChIP optimizations: purification of DNA products

Tue May 29, 2007

In a change for this round, I used both phenol chloroform and the Qiagen clean. For the phenol chloroform, I added 200 μl TE and 4 μl protease K as in Lee *et.al.* Nature Protocols. For the Qiagen kit, I just added 4 μl protease K. I incubated for 1 hr at 45 C.

2.4.7 second round ChIP optimizations: dynamoHS qPCR

Wed May 30, 2007

I switched to 10 μl rxns to save money and because I didn't have enough master mix to use 20 μl . I hope to go back and try the 20 μl rxns to see if it makes a difference. Since the total volume of master mix was much less, I put the master mix in PCR strips rather than a reservoir.

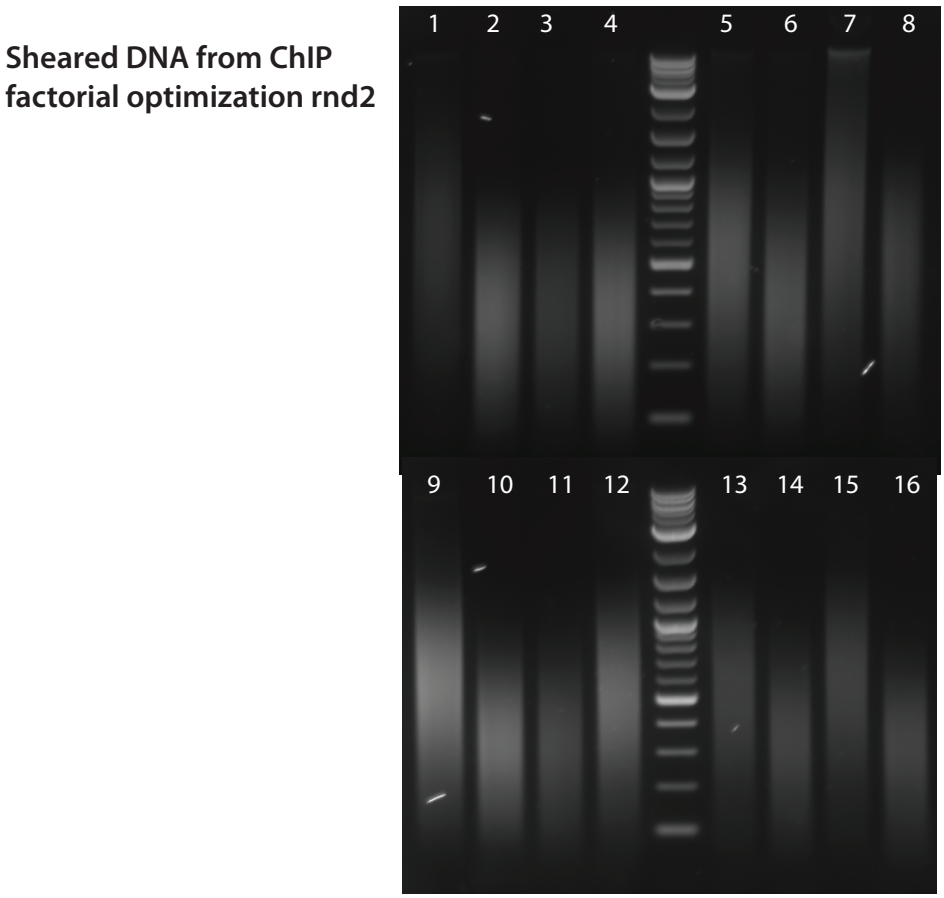


Figure 2.6: 1.5% gel run 45 min at 120V; 16 sheared randomized samples from factorial round 2

Plate 1: A302JWNE Plate 2: A302JWNF

Both of the plates had B in the bottom-left corner. HOWEVER, in the previous experiment I used B as the no-antibody sample, whereas in this experiment it was the antibody sample.

Here is the raw data for the hsDynamo factorial rnd2 samples 1-8 and factorial rnd 2 samples 9-16.

2.4.8 Summary of second round results

Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see factorialBatchRound2.m for the script).

Second round ChIP factorial results			
		hsDynamo qPCR	
id	*factor*	*effect*	*pvals*
1	amt starting DNA	-0.045	0.1259
2	formaldehyde	0.1516	0.0049
3	lyse	-0.2299	0.0013
4	shear	0.1599	0.0036
5		0.0173	0.7927
6		-0.0533	0.1818
7		0.0426	0.1094
8		-0.3404	0.0004
9		-0.0261	0.3011
10		0.0817	0.0382
11	antibody conc	0.3085	0.0006
12	bead conc	0.3593	0.0004
13	wash	-0.1931	0.0013
14	final DNA cleanup	-0.022	0.9478
15		-0.2114	0.0019

Brief Conclusions: Mon Jun 18 18:46:11 EDT 2007 I'm just going to summarize my gut feelings based on the above table.

Main effects (linear). If anything, lysing is not helpful (removing this will save at least an hour). Again formaldehyde is important. Here, it is not as strong an effect as last time. Perhaps 4% is a big overshoot. I definitely need to fine tune this concentration. More shearing was helpful alone this time, but did not seem to have an effect in combination with formaldehyde this time. I still think it would be useful to optimize these two together. More antibody was helpful. More beads was helpful. The normal washing was better than washing with TE alone.

What I'm a gonna do after round2.

1. use the original amount of starting DNA - increasing the amount change help anything, and slows things down because you have to grow more and use bigger tubes
2. back to 1% formaldehyde - I need to do a more in depth study of the formaldehyde concentration combined with changing the shearing amounts
3. don't lyse chemically - let the sonicator do it for you; saves a lot of time and some pipetting
4. shear fairly well - again needs to tried on its own with different formaldehyde concentrations; I want the minimal shearing time that will give the best results (since shearing is very time-consuming and labor-intensive)
5. antibody concentration middle - last time higher didn't matter. this time, lower did matter; I'm going to explore this further with a response surface
6. bead concentration - higher is better; will optimize with antibody conc in a response surface
7. don't wash too much - washing is slow, labor-intensive, and uses tons of tips; For now, I'm only going to do the two TE washes. I'm going to explore this latter in more detail. I realize this was a significant effect, but the effect size was not so large large that removal of the washes will cause the proceedure to fail

8. use Qiagen PCR purification for final DNA cleanup - had no effect; plus the results with the Qubit and the Qiagen purifications showed they were incredibly consistent down to very small amounts of DNA (see Figure 2.8).

2.5 Response surface 1: beads vs antibody

Given the results from factorial round 2, I decided to optimize the values of the antibody and bead using the method of steepest decent.

2.5.1 preparing the sheared chromatin

Wed Jun 13, 2007

I'm just going to prepare a batch of sheared DNA/chromatin to optimize the bead and antibody concentrations. I'm not going to prepare a separate sample for each concentration beads and antibody to try. Rather, one tube of sheared DNA is enough for 4 samples, so I'm just going to make 3 tubes of sheared DNA, which is enough to try 12 different samples.

The protocol is basically the ChIP protocol with a couple modifications using what I learned in round 2. The modifications are: stay with 1% formaldehyde, don't lyse with chemicals (put cells in Dilution buffer and let the sonication pop them open). I'm also going to do the washes with TE two times rather than using all of the salt buffers for reasons I mentioned above (see section 2.4.8 on page 126).

The sheared chromatin (4x30secx20%) for all three samples was run out on an agarose gel (Figure 2.7).

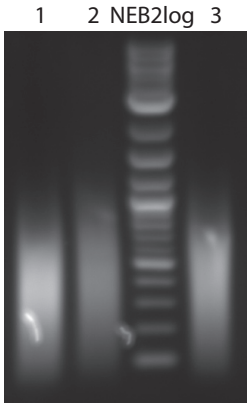


Figure 2.7: Sheared chromatin on agarose gel. This sheared DNA will be used for the response surface of beads/antibody.

Brief Conclusions: So far so, good. As it should be - I've sheared cells for ChIP about a million times now....

2.5.2 testing ability of Qiagen PCR prep to work with minute quantities of DNA

Thu Jun 14 19:42:30 EDT 2007

In factorial round 2, I found no significant difference between phenol:chloroform + EtOH vs Qiagen PCR purification. I wanted to dig into this further with the current batch of sheared chromatin, because the Qubit just arrived in the mail. The Qubit is a fluorometer from Invitrogen built just for quantifying DNA, RNA, and protein using Invitrogen's dyes. The dyes have a few important characteristics. First, the DNA dye is very specific for DNA *not* RNA, so you measure DNA concentrations almost independently of the amount of RNA in the tube (same holds in the other direction for the RNA dye). This specificity is a nice feature, because I can see if I still have RNA in my sample despite having put RNAse cocktail into my sample prior to sonication (e.g. maybe the sonicator deactivates the RNAse). The second key feature of their dyes is that they allow you to measure *very* small quantities of DNA (down to 10 pg/μl). I still don't know how much DNA I pull down with the ChIP proceedure, but I do know it is smaller than I can measure with the nanodrop (I've tried with the nanodrop and only get rubbish). The Qubit might allow me to quantify this DNA.

The goal for this second is to take the three sheared samples I described in the previous section and to dilute them with TE to several different concentrations. Then I'll clean them up and quantify with the nanodrop and the Qubit.

Qubit and Nanodrop readings from Qiagen purified sheared DNA					
Sample	**μl sheared DNA**	**Qubit (ng/μl)**	**Nanodrop (ng/μl)**	**Nanodrop (260/280)**	**Nanodrop (260/130)**
1	100	377	454.45	1.81	1.87
1a	50	158	168.28	1.82	1.81
1b	10	26.2	41.84	1.73	1.33
1c	5	14.6	29.52	1.67	1.05
1d	1	3.06	42.98	1.56	0.63
1e	0.5	1.174	12.35	1.42	0.71
1f	0.1	0.228	37.13	1.52	0.55
1g	0.01	0.0339			
2	100	297	356.3	1.8	1.78
2a	50	140	167.78	1.78	1.61
2b	10	23.4	44.55	1.72	1.19
2c	5	11.68	25.99	1.57	1.08
2d	1	3.02	15.58	1.4	0.71
2e	0.5	1.454	16.44	1.28	0.96
2f	0.1	0.316	26.69	1.48	0.63
2g	0.01	0.0347	13.09	1.37	0.72
3	100	303	333.13	1.81	1.82
3a	50	127	150.47	1.79	1.78
3b	10	21.6	44.56	1.73	1.06
3c	5	12.1	27.11	1.7	1.03
3d	1	1.974	12.44	1.65	0.8
3e	0.5	1.216	13.31	1.51	0.8
3f	0.1	0.252	12.72	1.47	0.66
3g	0.01	0.0312	12.78	1.41	0.69

The above table is easier to see as a graph (Figure 2.8).

Brief Conclusions: I certainly wasn't expecting my graph to look this good (Figure 2.8). Maybe if I'd just diluted a known concentration of DNA, I'd expect a graph this nice. But this is a Qiagen PCR Purification of different quantities of sonicated ChIP starting material. Not only is the Qubit

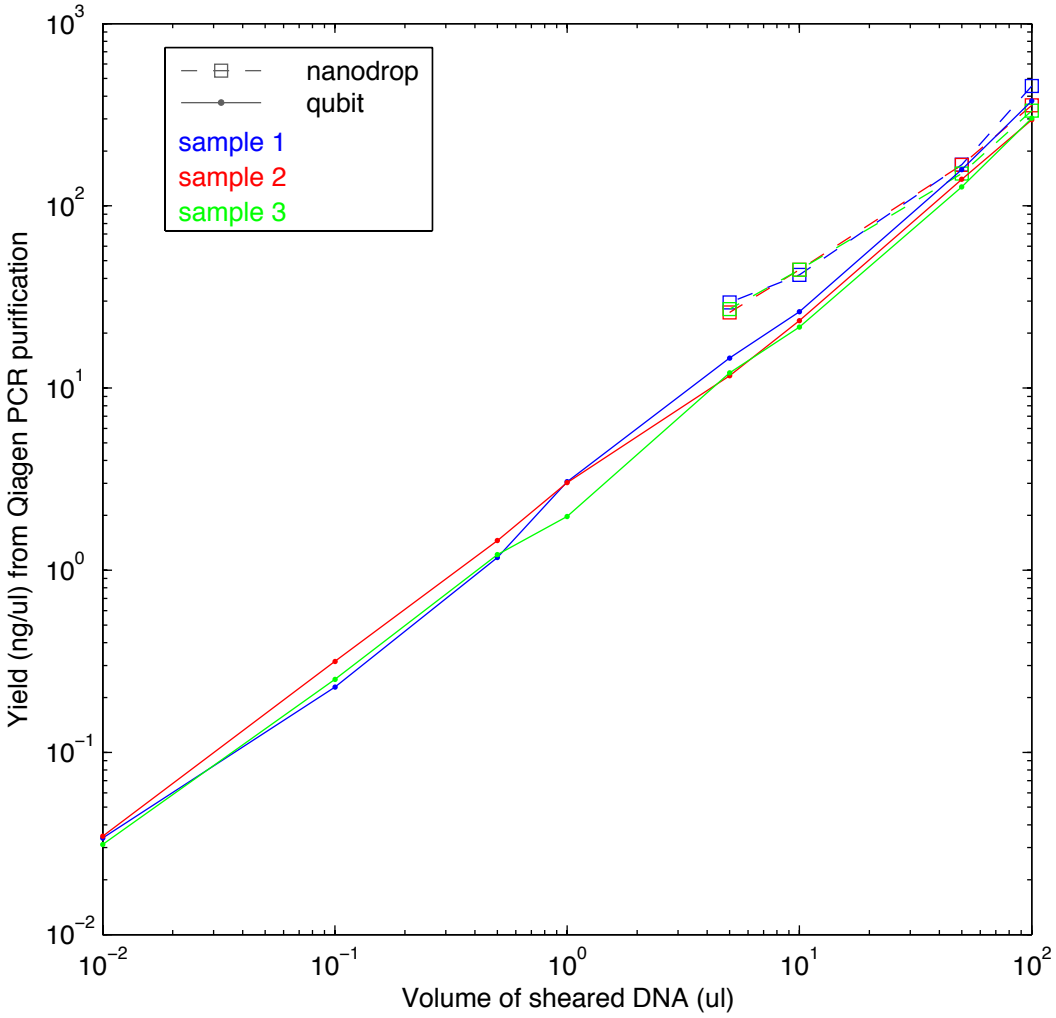


Figure 2.8: fill in

linear, the Qiagen cleanup must be extremely consistent across different concentrations of DNA as well. Notice that the nanodrop begins to overestimate the quantity of DNA at around 20-40 ng/μl . According to this figure, the nanodrop doesn't really work at all below 20 ng/μl . In general the nanodrop is about 10-20% higher then the Qubit readings. Perhaps this reflects the RNA remaining in the sample? Overall, this experiment went really well, and it increases my confidence to totally switch to Qiagen PCR purification kits for cleaning up the final ChIP DNA rather than slow, laborious, and hazardous phenol:chloroform extraction.

2.5.3 optimization of bead and antibody concentrations

Now that we've learned a little about what matters in the ChIP protocol, we can start to push forwards and optimize what matters. Our previous factorial experiments pointed out the factors and allowed us to speed up the protocol by removing things that didn't matter from the protocol. What we want to know now is can we get more enrichment. The first step is the optimization of bead and antibody concentrations, which were found to be important in factorial round 2.

The stuff I describe now comes from the books *Statistics for experimenters, 2nd Edition* pages 489-537 and from *BOX on QUALITY and DISCOVERY* pages 146-169.

Using least squares, I determined the parameters for altering bead and antibody concentrations using steepest ascent (antibody parameter = 0.1542; bead parameter = 0.1797). I rounded and scaled these parameters to 15 and 18 respectively and calculated four pairs of values for antibody and beads.

BEGIN NOTE

Ilaria figured out how to move right in the table in Figure 12.4 in *Statistics for Experimenters*. For the example in the book the second row is:

$x = \frac{-13}{28} \cdot 0.75 \cdot \frac{0.875}{0.875}$

the third row is:

$x = \frac{-8}{28} \cdot 0.75 \cdot \frac{0.375}{0.875}$

The first row is determined by the unit size chosen by the experimenter.

END NOTE

The values for antibody and beads determined by this method were:

antibody concentration (μg)	0.25	1.25	2.25	3.25
bead concentration (μl)	10	70	130	190

For each reaction, I used 100 μl of sample and 900 μl of dilution buffer. I ran two sets of these four concentrations. The main set was with dynal beads (the parameters actually came from the factorial analysis that used dynal beads only). The second set used agarose beads. I did this to give the agarose beads a second chance, as they performed a little better than the dynal in the first round. If they prove much better in this analysis, perhaps they're worth the extra hassle.

PCR rxns agarose vs dynal

I ran 10 μl PCR rxns for the dynal and the agarose beads. The average enrichment and standard deviation for each bead/antibody concentration set is shown in Figure 2.9. Bundling all of the genes into one mean and standard deviation creates a lot of noise (hence the big error bars), as some genes enrich much more than other genes.

Here is the raw qPCR data for the agarose and dynal rxns

Brief Conclusions: The bead/antibody concentrations were determined using the data from factorial round 2 where dynal beads were used. Therefore, it's a little unfair to compare the agarose and dynal beads at these concentrations, since the agarose might have had completely different concentrations. Nevertheless, dynal does seem to have less noise and is more consistent than the agarose beads.

The response curve for the dynal bead is quite nice, looks just like it does in the textbook. We can see how the enrichment increases as we pile on more beads and antibody. One thing we don't know from this plot (since I didn't carry it out far enough) is whether or not it saturates and if it saturates does the enrichment start to decline when more bead is added.

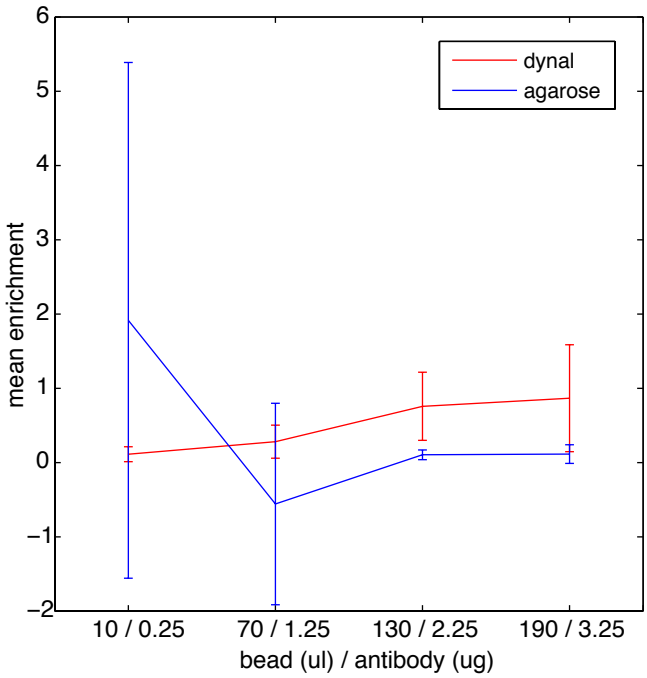


Figure 2.9: ChIP enrichment of positive control genes relative to random genes for bead/antibody concentrations determined via steepest assent.

PCR rxns old primers vs new primers

I ordered a new set of random primers and a new set of lrp primers for a couple reasons: 1) the previous primers I used were over a year old (though they have been in the freezer the whole time and only freeze-thawed 3x); 2) I want to be absolutely sure that what I'm seeing is not primer specific. In the new set of lrp primers, I picked a few of the primers from before and a few new ones. For the ones that were included from before, I redesigned a new primer for them (i.e. the primer will bind a slightly different location). The primer plate with the new random and new lrp primers was 2 μM as before and was laid out as follows:

Here is the raw qPCR data for the new lrp primers

Lrp factorial primer plate2												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	ynhG	apaG	ypdB	sanA	ybaO	arpB	infC	hybA	tdcA	mviM	ygfZ	ycfX
B	leuA	ilvH	lrp	dadA	oppA	aroA	livK	serA	ilvA	lysU	fimA	stpA

I tested the primers on sheared DNA, and ran them out on a gel (Figure 2.10). I think these gels are really only useful for making sure you didn't do something completely idiotic like mix the primers incorrectly. This check keeps you from wasting an entire qPCR plate, because none of the primers worked. However, it is important to do a second check once you have the qPCR results. On qPCR results for ChIP, genes that consistently have a Ct >= 30 when amplifying the immunoprecipitated DNA should probably be eliminated from further analysis. They just don't behave properly. Good Ct values will be from 20-28 or so.

The same enriched DNA was used for the new primers as had been used with the old lrp primers in the section above. The results comparing the new primers with the old primers are shown in Figure 2.11.

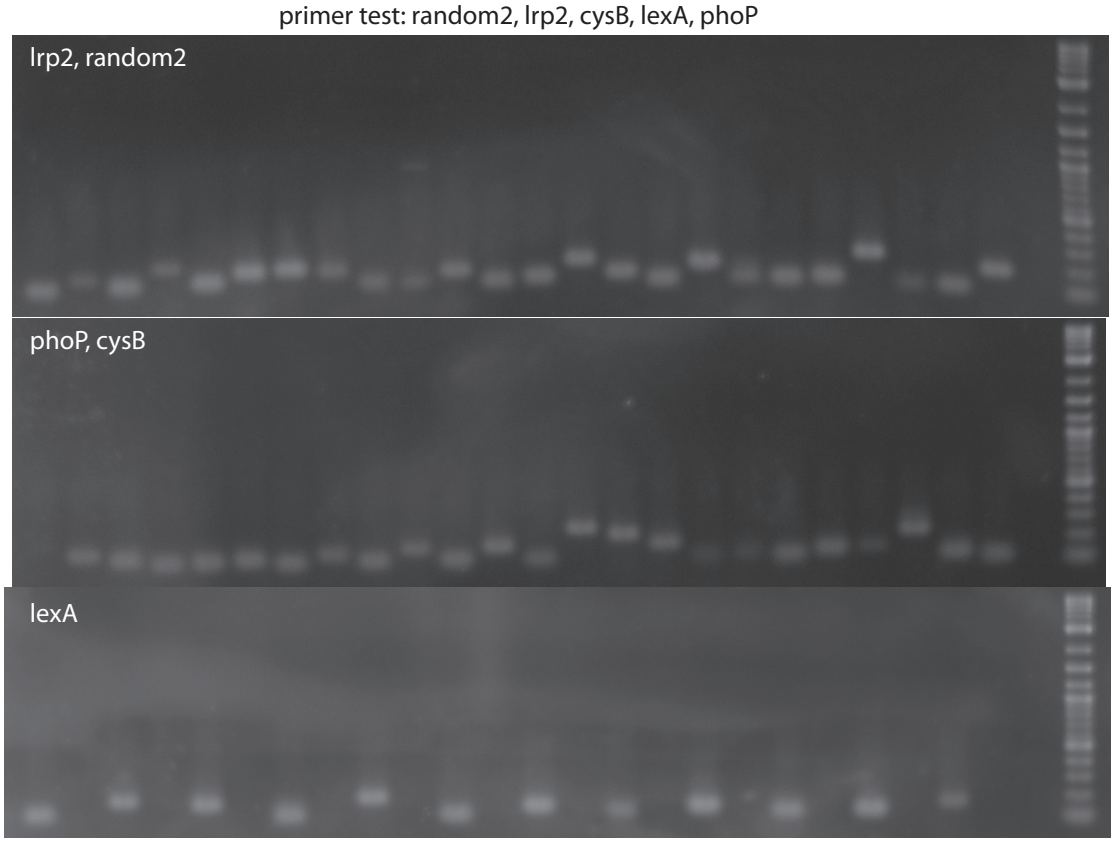


Figure 2.10: new lrp, random, cysB, and lexA primers amplified from sheared DNA

Brief Conclusions: The results are pretty dang similar between the two primer sets. Very nice... The only annoying thing was that very few of the primers I designed for lrp worked properly (i.e. the Ct values were too high, indicating that they weren't binding the DNA efficiently enough to be reliable). All 12 of the random genes worked beautifully though. Might be a good idea to make a plate3 that is a composite of 1 and 2. Would contain all N primers that I know work well: livK, serA, serC, pntA, dadA, yhjE, gtlB(?).

2.6 third round ChIP optimizations

Given the data in the antibody:bead response surface, I want to do one final factorial optimization of the second half of the ChIP protocol. The second half entails everything that occurs after the shearing.

For this final screening, I'm going to go with a full factorial design to allow easy interpretation of the results, particularly regarding any potential interactions we might find.

The factors are:

1. antibody (1.25 - 2.25 μg)
2. beads (70 - 130 μl)

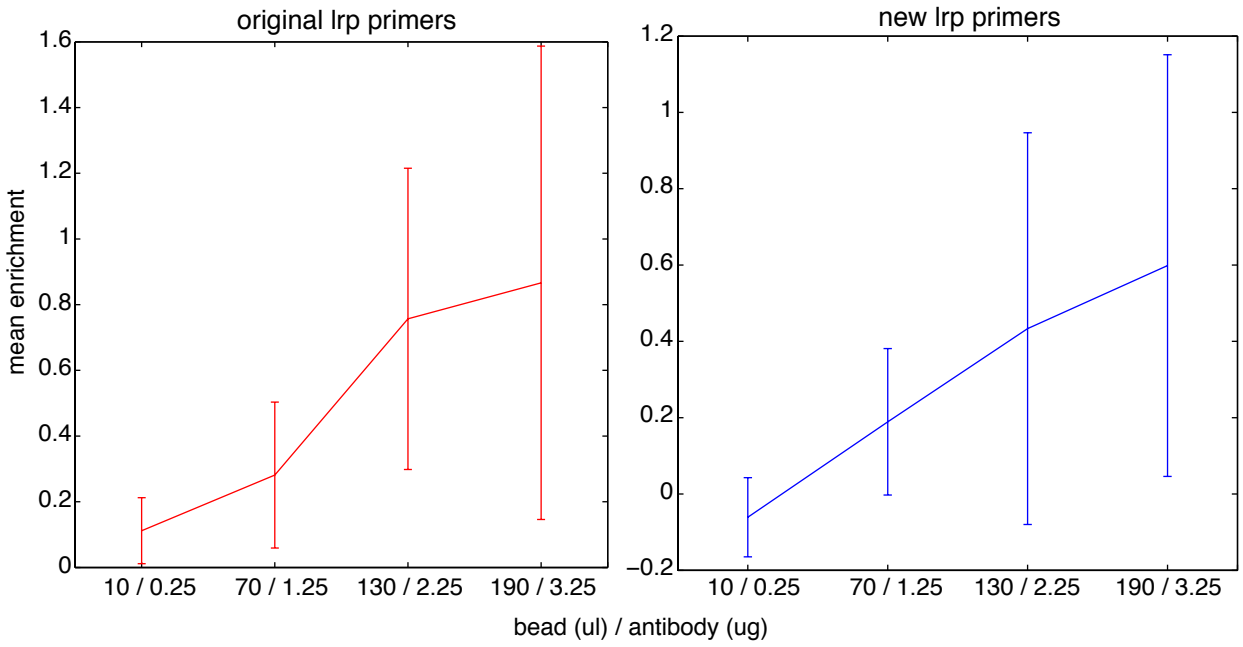


Figure 2.11: comparison of a newly synthesized set of lrp target primers relative to the performance with the old set I've been using up until now

3. silicon tubes (no - yes)
4. LiCl wash (no - yes)

The antibody and bead concentrations came from the previous response surface results above. Lot's of folks claim siliconized tubes work better to reduce noise, so I gave that a shot. And I also brought back the LiCl wash, which was significant in a previous round. I wanted to check it in this more rigorous full factorial setup to hopefully get a better gauge on the importance of the salt washes.

In order to look potentially fit a nonlinear surface to this data, I also included 3 replicates of the middle points, 2 samples below the factorial points, and 3 samples above the factorial points (see the table below to figure out what I'm talking about).

ChIP factorial 3rd round setup					
ID	**silicon**	**bead**	**anti**	**liclWash**	**randomized order**
1	no	70	1.25	no	8
2	yes	70	1.25	no	2
3	no	130	1.25	no	3
4	yes	130	1.25	no	23
5	no	70	2.25	no	22
6	yes	70	2.25	no	13
7	no	130	2.25	no	19
8	yes	130	2.25	no	1
9	no	70	1.25	yes	6
10	yes	70	1.25	yes	10
11	no	130	1.25	yes	9
12	yes	130	1.25	yes	15
13	no	70	2.25	yes	14
14	yes	70	2.25	yes	24
15	no	130	2.25	yes	16
16	yes	130	2.25	yes	21
17	yes	100	1.75	yes	20
18	yes	100	1.75	yes	5
19	yes	100	1.75	yes	17
20	yes	10	0.25	yes	18
21	yes	40	0.75	yes	11
22	yes	150	2.75	yes	12
23	yes	190	3.25	yes	4
24	yes	250	4.25	yes	7

2.6.1 third round ChIP optimizations: checking the sonication

Wed June 20, 2007

I grew up 6 samples in 30 ml of LB. Sonicated 4x30secx20%. No chemical lysis. I could spec the cultures, because the plate reader wasn't working.

Thur June 21, 2007

I cleaned and spec'd the six samples (volume was the typical 50 μl):

Sample	**DNA (ng/ul)**	**260/280**	**260/230**
1	428.7	1.80	1.78
2	454.9	1.80	1.84
3	457.2		
4	431.4	1.79	1.73
5	396.3	1.80	1.73
6	441.5	1.80	1.83

Yields were slightly higher than my normal readings (normally I get 200-300 ng/μl). I ran the sheared samples on a gel (much later Figure 2.12).

Fri June 22, 2007

I ran the binding and wash steps for all 24 samples detailed in the table above. The samples were run in the randomized order, pulling from each of the 6 sheared samples until their was no sample left (e.g. randomized samples 1-4 used sheared sample 1, 5-8 used sample 2, etc. . .). The samples were placed at 65C overnite to remove the crosslinks.

Sat June 23, 2007

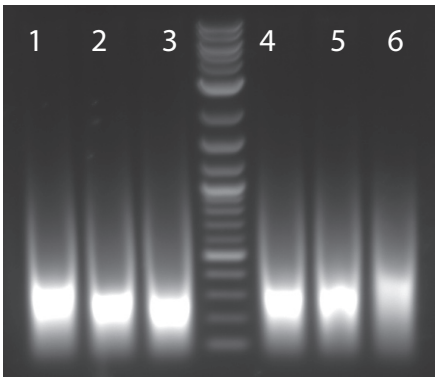


Figure 2.12: Six sheared samples on 1% agarose gel.

I cleaned up all 48 samples (24 samples without antibody, 24 samples with antibody) using Qiagen PCR cleanup columns. I eluted into 100 μl EB.

I ran 2 qPCR plates from on the purified DNA (this was the first 16 samples).

Mon June 25, 2007

I ran the final plate of the experiment.

Here is qPCR plate 1 of factorial 3, qPCR plate 2 of factorial 3, and qPCR plate 3 of factorial 3

Results from the 3rd round factorial

Figure 2.13.

Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see factorialBatchRound3.m the script).

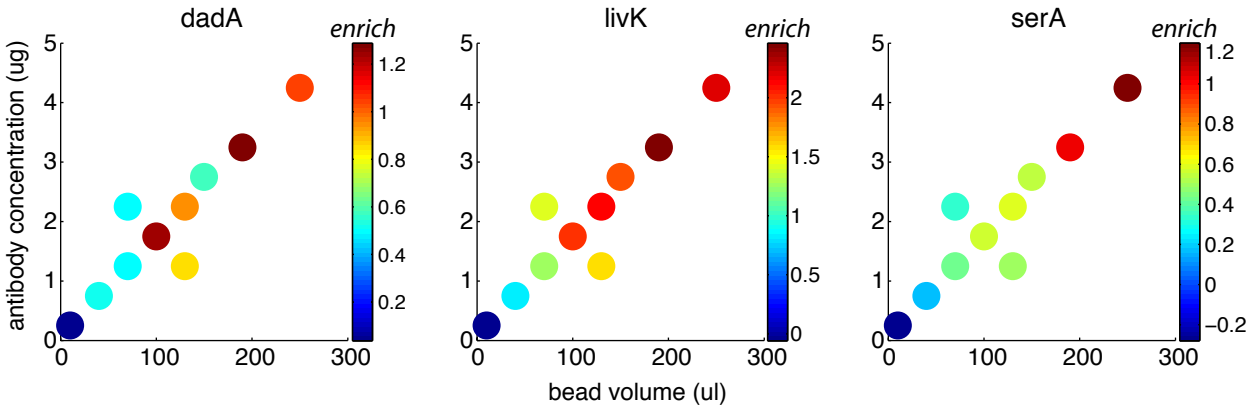


Figure 2.13: Enrichment (the z-axis displayed via color) certainly is dependent on antibody and bead concentrations. There appears to be some saturation (too much DNA?).

2.7 Preparations for future factors

I want to make a low-copy version of the AntiXpress-tagged lrp protein to see if bring the expression down even lower has an effect.

2.7.1 Cloning lrpB into a low-copy plasmid

To get the expression of LrpB down further, I want to change the copy number (which is currently very high) to low copy number.

```
Forward primer
----------------------------------------
     XhoI   promoter -35
ATAT CTCGAG TGTTGACAATTAATCATCCGGCTCGTAT

Reverse primer
----------------------------------------
     AvrII
TTAA CCTAGG ATTTGTCCTACTCAGGAGAGCGTTC
```

Tue Jun 19, 2007

I'm cloning the gene into the plasmid Ilaria uses (derived from the one U. Alon uses a lot). It has a very low-copy number (6 per cell?).

I amplified the lrp from the lrpB-TOPO plasmid that I've been using for ChIP. I added XhoI and AvrII using the primers above. I used 1 μl plasmid and 2 μl primer (from 10 μM stock) in the 30 μl reaction. I used the Finnizymes Phusion Taq and 30 cycles. Yields were:

Sample	DNA (ng/ul)	260/280	260/230
lrp-lowcopy	34.6	1.82	2.22

I ran 5 μl of the PCR rxn on a gel (Figure 2.14) to verify that it was the correct size (I was a little worried it would difficult to amplify over the 2' structure of the transcription terminator).

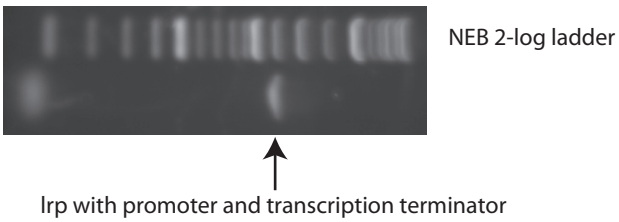


Figure 2.14: gel of lrp PCR prior to cloning into low-copy plasmid

I digested the DNA and the vector (obtained from Ilaria) using 2 μl NEBuffer2, 2 μl BSA, 15 μl template, 0.5 μl AvrII and 0.5 μl XhoI. I ran the digestions for 20 min followed by 10 min heat deactivation at 65C. I gel purified the digested vector using a Qiagen Gel Purification column, eluting into 30 μl . The gel was stained with Sybr Safe and I didn't take a picture of it, so that transformation efficiencies would be high (don't want the UV light messing up my DNA).

Unfortunately, I made 2 big mistakes with this and it didn't work the first time. First, I loaded my digested PCR into a gel, rather than cleaning it up with a PCR purification. I caught my mistake before I started the gel, so I sucked up the cut PCR product from the gel well with a pipettor. I cleaned up the DNA I sucked out of the gel well using a Qiagen PCR purification kit. I don't think this is an efficient way to clean up your DNA though :). Then in a bigger mistake, I assumed the plasmid was amp resistant and plated the transformation on an amp plate. The plasmid was actually kanomycin resistant. The negative control had zero colonies, but the positive control did have a few colonies (strange, but I'm not the first person to find this http://www.bio.net/bionet/mm/methods/2005-March/099322.html). I picked one of the colonies just for the hell-of-it, and it didn't grow in Kan.

Wed Jun 20, 2007

Since I only cut 15 μl of the PCR rxn yesterday, I had 15 μl left that I could cut today. I ran the same ligation as yesterday (but not for the vector, because I still had some gel-purified left from yesterday). I cleaned up the digestion using a Qiagen PCR purification kit, eluting into 30 μl . Then I ran the following ligation: 5 μl PCR digestion, 8 μl cut vector, 2 μl T4 buffer, 1 μl T4 ligase, 4 μl H_2O . I incubated 10 min at RT; I did not heat deactivate. Rather, I directly transformed 2 μl into DH5α competent cells.

2.8 forth round ChIP optimizations: chromatin concentration

The previous two rounds clearly show that bead and antibody concentrations are important. With more beads and more antibody and you get more enrichment. It's not clear what the surface of displaying the interaction between these two variables will look like. However, in the previous round I appears that I need a heck of a lot of beads and antibody to get the best enrichment. I want to fit this surface describing the interaction between beads and antibody, but it would be expensive to do given the current amount of beads that my results suggest I should use.

The question then becomes *are we saturating the beads and antibody with DNA*? If so, we should be able to drop the DNA concentration and move the quantities of bead and antibody down accordingly. If this is the case, then the ChIP protocol becomes much more flexible. Knowing the interaction between these three, you can always maximize enrichment by choosing the optimum concentrations of beads and antibody for each amount of DNA. By increasing the amount of DNA, you can (presumably) grab more DNA which might be better for cloning the enriched DNA. By decreasing the amount of sheared chromatin DNA, you can lower costs, particularly if you are doing sensitive qPCR reactions to verify enrichment.

2.8.1 the plan

To test this, I'm running the following full factorial with three factors.

ChIP factorial 4th round setup				
ID	**bead (μl)**	**anti (μg)**	**sheared DNA (μl)**	**randomized order**
6	130	0.75	100	1
7	40	2.25	100	2
2	130	0.75	25	3
8	130	2.25	100	4
5	40	0.75	100	5
3	40	2.25	25	6
4	130	2.25	25	7
1	130	0.75	25	8

2.8.2 grow, lyse, shear

Mon Jul 9, 2007

I grew 6 samples as in the previous factorial experiment. I grew them to 0.5 OD600. I lysed them via sonication.

Tues Jul 10, 2007

I cleaned and spec'd the six samples (volume was the typical 50 μl):

Sample	**DNA (ng/ul)**	**260/280**	**260/230**
1	207.8	1.84	1.93
2	226.7	1.83	1.84
3	212.3	1.84	1.94
4	235.4	1.83	1.92
5	198.6	1.84	1.90
6	192.5	1.84	1.95

2.8.3 immunoprecipitation

Tues Jul 10, 2007

I proceeded according to the factorial setup above using the randomized order. I used 900 μl of dilution buffer for all rxns whether I used 25 μl of sheared DNA or 100 μl of sheared DNA. Since there were only 8 conditions, I only used samples 1 and 2 of sheared DNA (immunoprecipitations with randomIds 1-4 used sheared sample 1 and 5-8 used sheared sample 2).

2.8.4 elute DNA and qPCR

Wed Jul 11, 2007

I cleaned up the samples with a Qiagen PCR purification kit. Unfortunately, I dropped a bunch of the tubes onto my bench as I was closing the tubes for the final time. The only samples that noticibly lost a little volume were 2c and 7c. However, since I'm looking at relative ratios, it shouldn't make a big deal.

I ran 10 μl qPCR using hsDynamo and the lrp plate 2 with random primer set 2.

Here is the raw data for the qPCR rxns factorial rnd4 all 8 samples.

2.8.5 Summary of fourth round results

Click for a little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see factorialBatchRound4.m).

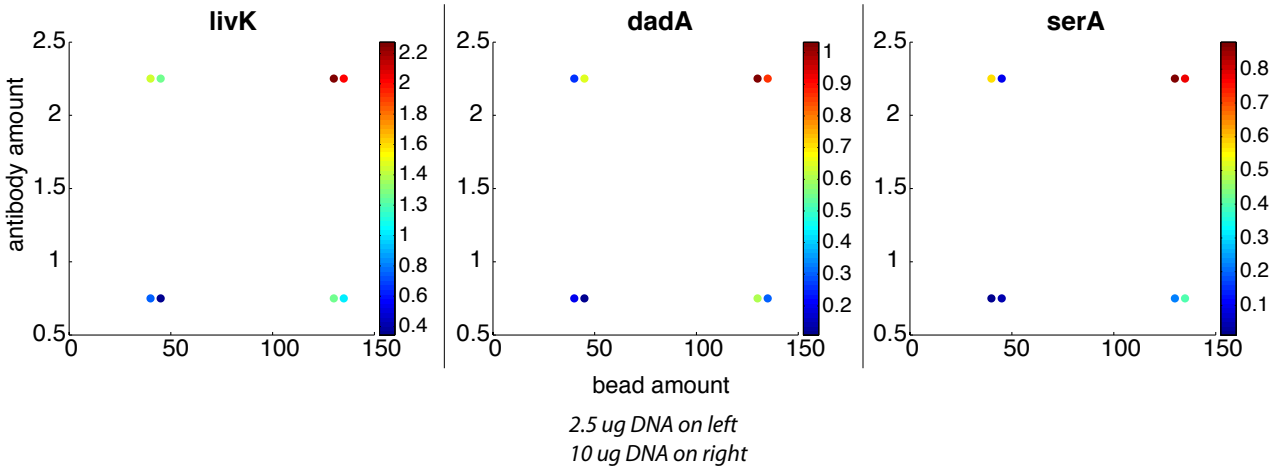


Figure 2.15: figure summarizing results of the 4 factorial experiment. Note that I've shown the forth dimension (DNA concentration) by putting the points right next to each other. The low DNA concentration is on the left and the high DNA concentration is on the right. This makes it appear that the low DNA sample has slightly less beads than the high DNA sample, this isn't the case. I just change the amounts slightly on purpose so the datapoints wouldn't overlap. For an alternative view see Figure 2.16

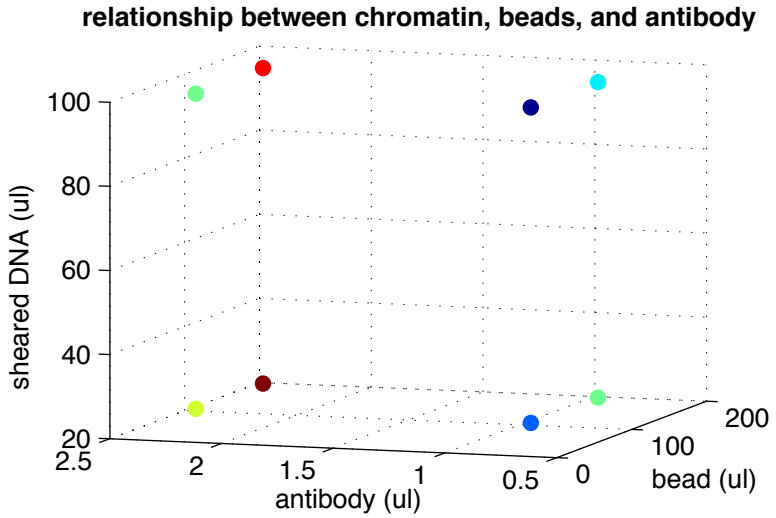


Figure 2.16: figure summarizing results of the 4 factorial experiment. this figure is plotted in 3d for an alternative view to Figure 2.15 for livK.

Brief Conclusions: Very nice when your hypothesis turns out to be right. In Figure 2.15 notice that all of the points on the left (low DNA) are more enriched than those on the right, and in Figure 2.16 the top points (top as in the points on the top if you image this thing is a cube; that is DNA conc = 100 μl) are pretty much always less enriched than the corresponding point right below them (i.e. DNA = 25 μl).

So it does look like we'll be able to scale the amount of antibody and bead down but using much smaller amounts of sheared chromatin starting out. I don't know yet how this will be affected by going to low copy number plasmid (i.e. will I need more sheared chromatin?). I also still need to see the affect of bead and antibody saturation (i.e. do things start to go bad when the sample is overwhelmed with beads and antibody?

Since I have 4 sheared DNA sample left, I think I want to drop the concentration even lower. Where I stand now, I should use 25 μl of sheared chromatin (appx 2.5 μg of DNA) with 100 μl of beads and 2 μl of antibody. I'd like to get the optimum bead amount down to 50 μl and then fill out the entire matrix of antibody to bead.

2.9 fifth round ChIP optimizations: finishing up the bead antibody surface

Thu Sep 20 13:21:05 EDT 2007

After a long break (for postdoc interviews and to publish a couple computational papers), it's time to finish this thing up. Given the recent confirmation that our chromatin was saturating the beads and the antibody, we can tone things down and figure out the surface in more detail without using 10 gallons of expensive dynal beads.

2.9.1 fifth round (bead/antibody surface) experiments

Wed Sep 19, 2007

I grew 50 ml lrpB in one 250 ml flask (originally I wanted to keep with my standard 25 ml in 125 ml flask, but all of the flasks were dirty). I grew the cells from a 1:100 dilution for 3 hr 30 minutes to an OD600 of 0.456; I added 1% formaldehyde, incubated 10 min, and quenched with glycine. Sonication was 4x30secx20%. 100 μl of the sheared chromatin was left overnite at 65C to reverse crosslinks.

Thur Sep 20, 2007

I cleaned up the crosslinked-reversed DNA with a Qiagen PCR purification kit. I eluted into 50 μl of EB, yields were:

Sample	DNA (ng/ul)	260/280	260/230
Sample 1	169.9	1.83	1.74
Sample 2	188.1	1.83	1.65
Sample 3	231.4	1.82	1.66

I ran 2.5 μl of all three samples on a 1% agarose gel (Figure 2.17).

Fri Sep 21, 2007

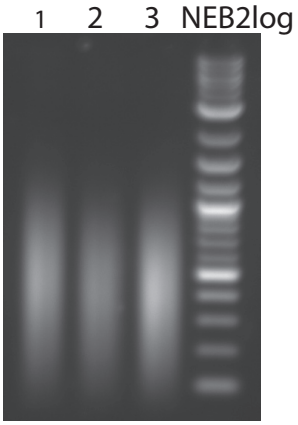


Figure 2.17: Sheared chromatin for ChIP. Average length is around 400 bp.

Based on the previous rounds and a little intuition, I decided generate the antibody/bead surface with 16 samples in a 4x4 matrix of antibody (0.75, 1.5, 2.25, 4.5 μl) and beads (40, 70, 100, 200 μl). As always, I randomized the order of the samples excel file of table:

sampleID	randID	antibody (μl)	bead (μl)
1	16	0.75	40
2	6	1.5	40
3	14	2.25	40
4	12	4.5	40
5	1	0.75	70
6	13	1.5	70
7	5	2.25	70
8	9	4.5	70
9	15	0.75	100
10	11	1.5	100
11	10	2.25	100
12	3	4.5	100
13	7	0.75	200
14	2	1.5	200
15	8	2.25	200
16	4	4.5	200

I used 17.5 μl of chromatin sample 3 (231.4 ng/μl = 2.025 μg /sample) for each precipitation. The preciptations were run with a 30 min antibody incubation and a 30 min bead incubation. The beads were washed 2x and eluted into 210 μl Elution buffer (from the Lee et al. Nature Protocols paper).

I also calculated the amount of DNA from the previous rounds of ChIP where I varied the antibody. When I got sloppy earlier, because I didn't think chromatin concentration mattered too much, I began just always using 100 μl of chromatin. Now that I know that chromatin concentration affects the saturation point of the beads and the antibody, I can actually use bead/antibody data at those different chromatin concentrations to include chromatin concentration as a third variable in my model (nice when sloppiness is actually useful). I wrote a matlab script to get the enrichment values along with their corrresponding chromatin, antibody, and bead concentrations (see the function combineDNA_Anti_Bead for the script). Note including the previous run which is at around 2 μg per sample, I've used 2.65, 10.6, 16.3, and 21.2 μg of chromatin in different runs). For most of these samples, I actually pulled from one of several sheared chromatin samples for each precipitation reaction. I just took the average chromatin concentration from all of the sheared chromatin to

calculate the amount for the experiment, since I didn't actually specify in the experiment which of the chromatin samples I used for each reaction.

Mon Oct 1, 2007

There's been a bit of a delay in running the qPCR on those bead/antibody samples, because our 12-channel small volume pipettor was broken. Turns out it was irrepairably broken, so I ordered two new ones from biohit a mechanical (m10) and an electronic (proline). The proline arrived today, so I'm going to run the first plate and give the new pipettor a try. I to make an lrp-superplate containing all of the positive control genes that I'd used that I knew worked (i.e. the primer pair got a decent Ct value in the mid-to-low twenties). I used the random_primer_plate2 primers for the random genes (just as I've been doing for most of these experiments). I randomly ordered the 12 positive primers, except that I left the positives from the previous bead/antibody runs in the same location on the plate (just in case there was something magical about the locations they were in). I also included the serA1kb primer which is 1kb away from the known serA site. And I included 3 replicates of the livK from the second plate, so that for one gene I'd have a triplicate qPCR technical replicate for all of the ChIP samples. The primers with a "2" after the gene name are from the second primer plate I ordered from IDT.

Lrp factorial superplate (2 μM)												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	ynhG	apaG	ypdB	sanA	ybaO	arpB	infC	hybA	tdcA	mviM	ygfZ	ycfX
B	livK	serC	serA1kb	dadA2	livK2b	pntA	livK2	serA2	yhjE	gtlB	serA	livK2c

I ran samples with random ids from 1-8 on the first plate. Unfortunately, either the new proline pipettor doesn't seal well on the tips or I didn't give the pipettor enough time to charge before I started, because by the end the pipettor was aspirating very unevenly if at all. Because of that, there were a few wells on my 384-well plate that certainly didn't have the correct volume. So I think I'm going to repeat these qPCR experiments when the mechanical m10 pipettor arrives and hope for the best.

Here is the raw data for the qPCR run bead antibody samples 1-8 proline pipettor and the matlab script I used to partially analyze them is here: factorialBeadAntiSurface1, though I didn't do too much analysis because I wasn't happy with the quality of the data. From my quick scan, the general trend of the data looked about right.

The data for livK were by far the best because, as you can see in the primer table above, I placed three technical replicates for the livK samples (livK2, livK2b, livK2c).

Wed Oct 3, 2007

I received a new low volume pipettor an mLine (m10) mechanical pipettor. I decided to use the to test samples 9-16 of this antibody-bead surface. I'm skipping trying to clean up samples 1-8 for the moment, because I only have enough of 1-8 to give it one more shot. So I'd prefer to try the new pipettor on samples 9-16 where I have to chances left. Learning from the success of including technical replicates on my plate. I abandoned the lrp-super primer plate from the last experiment and created yet another primer plate. This primer plate contained only dadA, serA, livK, and pntA each in triplicate. dadA, serA, and livK were all from the second lrp plate I had synthesized by IDT. The column ordering was randomized with the matlab randperm() function:

Lrp factorial technical replicate plate (2 μM)												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	ynhG	apaG	ypdB	sanA	ybaO	arpB	infC	hybA	tdcA	mviM	ygfZ	ycfX
B	dadA	livK	serA	serA	dadA	pntA	dadA	pntA	serA	livK	pntA	livK

The new pipettor proved to be a godsend, and thankfully, I don't think I'll have to worry too much about poor tip seal on my low volume multichannel. The pipetting was quite accurate. I liked the pipettor enough to write a review of the m10 on IzziD. I'll delay discussion of the qPCR results until after I've run samples 1-8 using the new mLine multichannel.

Wed Oct 4, 2007

I ran samples 1-8 using the m10 pipettor. No problems.

2.9.2 Summary of fifth round (bead/antibody surface) results

Fri Oct 5, 2007

Click for a little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see factorialBeadAntiSurface2.m).

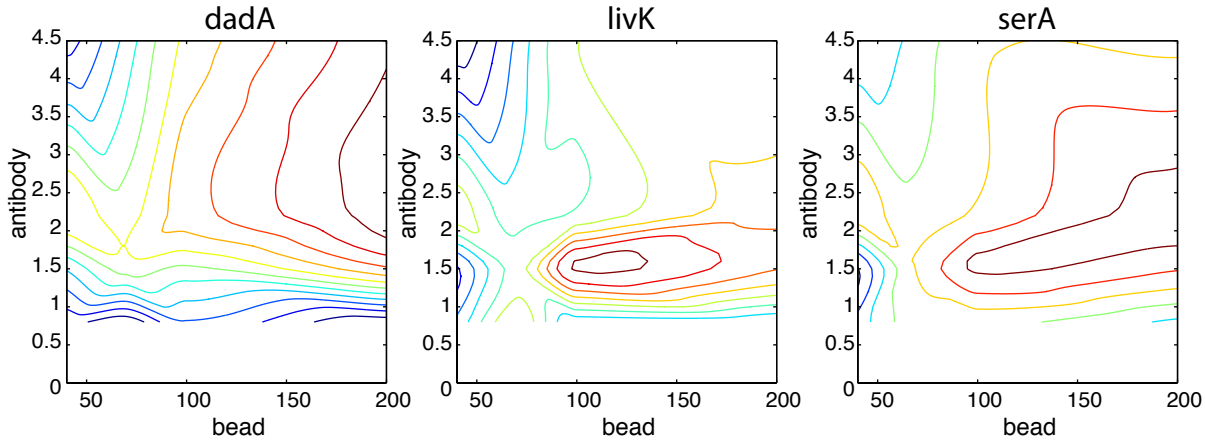


Figure 2.18: Contour plots from round 5

Brief Conclusions: Figure 2.18 shows contour plots fit using the median qPCR technical replicate enrichment value for dadA, livK, and serA. Note that in creating this figure, I cheated and removed one point (bead = 70 μl , antibody = 4.5 μl) that was vastly more enriched than other samples around it. I'm guessing this was some kinda outlier. I'd need a few more replicates at that point before I'd believe that 70 μl of bead with 4.5 μl of antibody mysteriously does better than slight deviations from those values. You can see from the contour plots that there seems to be a narrow window for livK and serA where you get optimal enrichment. I think this is due more to my sampling. I was expecting (hoping?) to see that at the extreme upper-right corning the values would taper off a little like they do (for example from some sorta bead or antibody saturation). However, I only have antibody samples at 2.25 and 4.5 and I think the small size of the bright red region is due to the cubic fit not having any samples from 2.25 to 4.5. To clean up this surface, I'd need to fill in a few of the holes, and hopefully the bright red section would increase. I just realized I forgot to add a colorbar to these charts. But really I just want to show the trend, rather than the absolute enrichment anyways.

So to conclude, I think that the surface – particularly for livK and serA – looks pretty promising and as I expected it to look. The only negative thing, I dropped the surface to 2 μg of DNA based on

the experiment above (see section 2.4.8 page 125) where I showed I could get more enrichment with less bead and antibody by using less chromatin DNA (presumably the chromatin was saturating my bead and antibody). However, the maximal regions in Figure 2.18 don't require *that* much less bead or antibody than the maximal values in Figure 2.13 where I used much more DNA.

I was hoping to use the combined bead/antibody/DNA values to fit a nice model showing the relation ship between these three variables using a 2nd order taylor series model (combineDNA_Anti_Bead.m is the matlab script to at least grab all of the values into one set of variables). So far the relationship isn't really making sense. It works at one level of DNA concentration, but really looks weird as I make 3d surfaces at different DNA concentrations. Ilaria set up all of the modeling stuff. I think the problem is overparameterization given the sparsity of antibody bead concentrations tested at all DNA concentrations besided the 2 μg used in this round 5.

2.10 Sanity check: again, does lowing DNA concentration really allow us to lower the amount of bead and antibody

Fri Oct 5, 2007

After section 2.4.8, it seemed clear that I'd be able to drop down the DNA concentration lower and lower until M μl bead and N μl antibody would allow maximal enrichment (that is I could lower the DNA and likewise lower the amount of antibody and bead need to get maximal enrichment of that amount of DNA. all I needed to do was to figure out the mathematical relationship between the three variables). However after finishing section 2.9.2, I'm having my doubts (see the brief conclusions in that section for my reasoning).

So I'm going to look at DNA concentration for a second time. In this experiment, I'm going to try a wide-range of DNA concentrations to see if my hypothesis from section 2.4.8 strengthens or weakens.

I'm going to use four chromatin DNA concentrations (10 μg , 2 μg , 0.4 μg , and 0.08 μg) across two combinations of bead/antibody (50 μl /1 μl and 100 μl /2 μl). I used sample 2 (188.1 ng/μl) from section 2.9.1 as the chromatin sample. The randomized experimental design was (excel file):

sampleId	**randID**	**bead**	**antibody**	**DNA (μg)**
6	1	2	100	0.4
7	2	1	50	0.08
2	3	2	100	10
8	4	2	100	0.08
5	5	1	50	0.4
3	6	1	50	2
4	7	2	100	2
1	8	1	50	10

Sat Oct 6, 2007

I cleaned up the reversed-crosslinked enriched DNA using Qiagen PCR purifications. I used the lrp-technical-plate from round 5. However, in that plate one of the dadA replicates didn't work (presumably I forgot to put the primers in that particular colume). I added primer from my stock to that broken well so that dadA would have all three of its samples for calculating it's median qPCR-based enrichment value.

Click for a little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see DNA_conc_data.m).

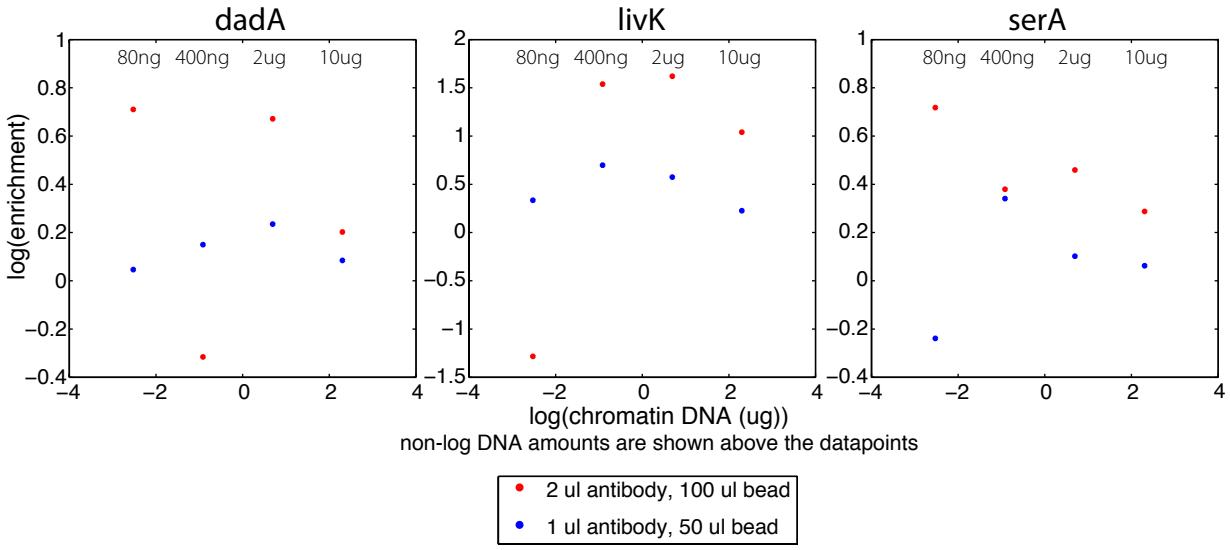


Figure 2.19: Sheared chromatin for four DNA concentrations using two combinations of bead and antibody

Brief Conclusions: The highest two concentrations in Figure 2.19 are the same DNA concentrations I used for the data in Figures 2.15 and 2.16. The two experiments are in agreement. I get better enrichment with 2 μg of DNA than I do with 10 μg . However, from Figures 2.15 and 2.16, I assumed the hypothesis that the beads and antibody were being saturated by too much DNA had become more believable. Using that hypothesis, I extrapololated that I should be able to keep dropping the amount of DNA until I could get a maximal enrichment with a smaller amount of bead and antibody (to drastically cut the costs of ChIP). The new results in Figure 2.19 don't support this extrapolation at all. After I get below 2 μg , the results become very inconsistent across my genes. With my extrapolation, you'd also think that the red and blue points (high bead/antibody -vs- low) would start to converge and eventually intersect; the blue points would be on top at low DNA and the red points would be on top at high DNA. Instead 100 μl beat / 2 μl antibody is pretty much *always* better than the lower concentrations.

Either there is some minimal threshold amount of bead/antibody or else this idea of DNA saturation is not completely true (or maybe not true at all?). If the saturation idea is not true at all, I don't have any way to explain why the 2 μg has now repeatedly out performed 10 μg of DNA.

2.11 Does the volume of the precipitation matter?

Mon Oct 8, 2007

Given the previous results suggesting that lowering DNA chromatin concentration wouldn't allow me to likewise use less bead and antibody and still obtain maximal enrichment (section 2.10), I decided to try one final thing to see if I can reach my goal of being able to control bead and antibody concentrations. Perhaps the reason more bead works better is because the beads then represent a

larger precentage of the volume in my tube. I usually use 1 ml in each precipitation reaction. With 200 μl of beads, 20% of my reaction is beads, but with 50 μl of beads, only 5% of my reaction is beads. The manual that Invitrogen sends with their beads also suggests minimizing your volume. However, all of the papers I've read doing ChIP with dynal beads use at least the volume I use – often more.

Similar to the previous DNA concentration test (section 2.10), I'm going to try four levels of my feature of interest using both 1 μl :50 μl and 2 μl :100 μl of bead:antibody respectively. The randomized experimental design is (excel file of experimental design):

-	randID	sheared chromatin (2 ug)	dilution buffer	antibody	bead
5	1	24	151	1	50
8	2	24	13.5	2	100
7	3	24	13.5	1	50
3	4	24	426	1	50
4	5	24	426	2	100
2	6	24	976	2	100
6	7	24	151	2	100
1	8	24	976	1	50

Note that when I prepared the beads, I resuspended them all (both 50 μl and 100 μl beads) into 100 μl of PBS. So the final volume in each tube is the volume of dilution buffer plus 24 μl of sheared chromatin plus 100 μl of bead/PBS. For the two TE washes for each sample, I used the volume of TE corresponding to dilution buffer plus 100 μl (e.g. for sample *randID=1* I used 251 μl of TE for each wash; for *randID=2* I used 113.5 μl TE for each wash).

Tues Oct 9, 2007

I qPCR'd the 8 samples.

Click for a little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see `volume_bead_antibody.m`).

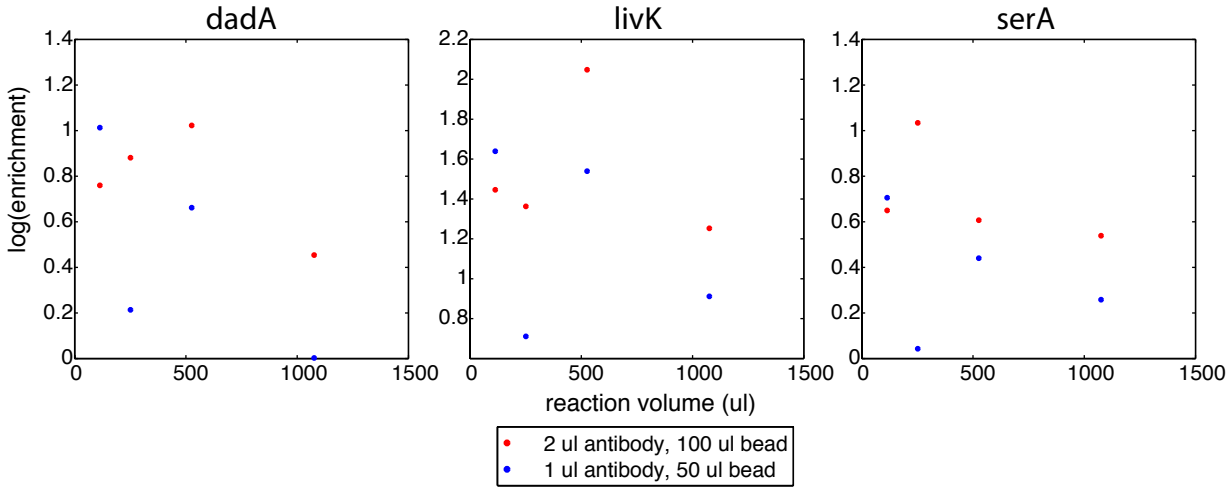


Figure 2.20: Enrichment when using a series of different volumes for the precipitation reaction

Brief Conclusions: Initially when I saw the resulting enrichment -vs- volume plot, I thought I just had noise and that volume didn't matter (Figure 2.20). However, if you look ignor the second point from the left (volume = 251 μl). The low bead/antibody samples (blue) follow a pretty darn straight line that maxes out at the most concentrated volume. The high bead/antibody samples (red) aren't as clean. They initially do better as you concentrate them more, but then they perform worse as you concentrate them further. Above (section 2.10), when I was scanning the DNA concentrations, I assumed that at some point the red and blue samples should intersect with the blue becoming more enriched – perhaps that's what I'm seeing here? However, I haven't forgotten that the second point from the left might not be an outlyer at all and all I'm hoping for here could just be a bunch of rubbish. I think there is a strong enough hint of a signal in this data to inspire me to give this a second try to replicate. Unfortunately, I'm outta sample, so first I need to prepare some sheared chromatin...

2.12 Sanity check: does the volume of the precipitation matter?

Sun Oct 14, 2007

My results in section 2.11 were decent enough that I want to explore further if the reaction volume is important for ChIP. Unfortunately, I'm outta fresh sheared chromatin. So I just started an LrpB (i.e. AntiXpress-tagged Lrp) overnite culture.

Mon Oct 15, 2007

I grew up 3x50 ml culture in 250 mml baffled flasks using a 1:100 dilution of the overnite. After one-hour I added 0.01 μM of IPTG. After 3 hr and 30 minutes, the ODs for the three cultures were in the correct range: 0.53, 0.518, and 0.485; I took two 15 ml samples from each flask and added the standard 1% formaldehyde for 10 min followed by quenching with glycine and two washes in ice-cold PBS.

Mon Oct 16, 2007

The six samples were sonicated 4x20%x30sec. 100 μl sample was crosslink-reversed and purified with a Qiagen PCR purification column. 5 μl of each sample was run on an agarose to verify the correct shearing range (Figure 2.21). The yields (elution into 50 μl) were:

Sample	DNA (ng/ul)	260/280	260/230
1	127.5	1.86	2.00
2	139.9	1.85	1.96
3	126.4	1.84	1.95
4	125.4	1.78	1.37
5	129.5	1.79	1.43
6	121.2	1.84	1.93

Brief Conclusions: As with my formaldehyde/shearing samples (section 2.15 page 153), the yields were about half of my usual values. The only thing I can think of is that I believe I recieved a new tube of RNAse Cocktail [Ambion] prior to the formaldhyde/shearing samples. Perhaps, the old cocktail was getting old and didn't remove all of the RNA?

Tues Oct 16, 2007

I ran a randomized enrichment volume scan using the exact values as the previous attempt (section 2.11), but the order was rerandomized. The design was (excel file of design):

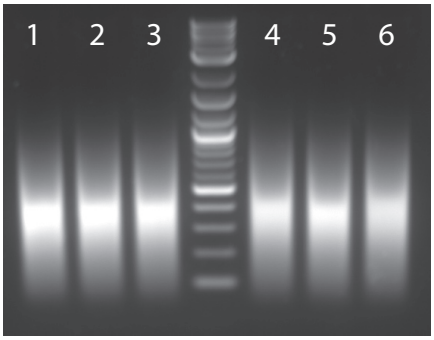


Figure 2.21: Six samples where sheared, crosslink reversed, and cleaned up. 5 μl of each (appx 600 ng) were run in each lane. The average shearing length is around the expected 300-400 bp range.

-	randID	sheared chromatin (2 ug)	dilution buffer	antibody	bead
6	1	31	151	2	100
7	2	31	13.5	1	50
2	3	31	976	2	100
8	4	31	13.5	2	100
5	5	31	151	1	50
1	6	31	976	1	50
3	7	31	426	1	50
4	8	31	426	2	100

For each sample I used 31 μl of sheared chromatin sample 1 (127.5 ng/μl) from the table above.

Wed Oct 17, 2007

I cleaned up the crosslink-reversed samples using Qiagen PCR purification columns. I ran a qPCR on all the samples.

Click for a little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see `volume_bead_antibody.m`). This data was combined with the previous replicate (section 2.11).

Brief Conclusions: The faint signal hinting at a linear increase in enrichment with a decrease in the rxn volume is much cleared when the data from this new replicate is averaged with the previous replicate (compare Figure 2.20 and with the averaged data in Figure 2.22). For dadA and serA there is a clear linear increase in the enrichment. The livK behaviour is a little different; it appears to increase in enrichment followed by a small decay as the volume reaches the smallest levels (the pink points are noisy, but I think the cyan points are a good representitive of what I'm refering to – it's almost like a candy cane shape). The other gene I'm following (pntA) that isn't in this plot, also displays a similar shape to livK.

Unfortunately, the points at 251 μl only have one replicate, because the other sample for each of the two volumes was a pretty extreme outlier (it was >2x the log distance away). I threw out these outliers, but the 251 μl point remains quite noisy and I'd really like to clean it up.

Thur Oct 18, 2007

In an effort to try and clean up the 251 μl points, I ran three replicates at that point for both the 1:50 and the 2:100 antibody:bead combinations. The design was (excel file of design):

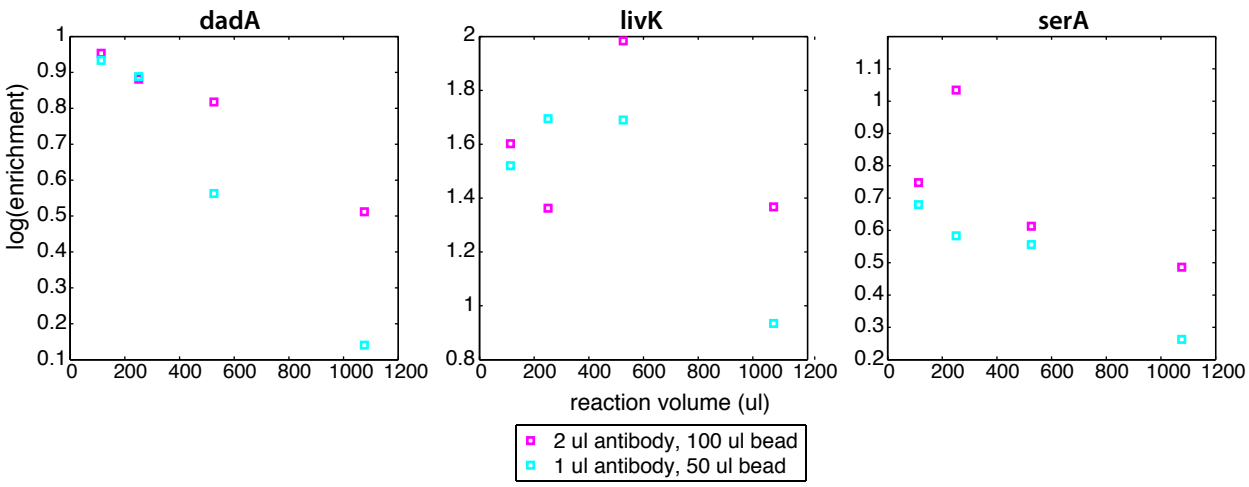


Figure 2.22: Combined data from the two volume scan replicates. All points are the average of two replicates *except* the two points at 251 μl . These each (i.e. 1:50 and 2:100 bead:antibody) had a screwy outlier (>2x log units away from all other points) that I removed.

id	randID	antibody	bead	volume
3	1	1	50	251
2	2	2	100	251
1	3	1	50	251
6	4	2	100	251
4	5	2	100	251
5	6	1	50	251

I cleaned up the enriched DNA with a Qiagen PCR purification column.

Sat Oct 20, 2007

I ran the qPCR plate, but unfortunately, the data was just a mess. Lots of reactions failed and those that worked had much higher Ct values than I'm used to. Later, I discovered that the primer plate was almost empty, so perhaps I wasn't pipetting my primer concentrations accurately. Either way, this data is not terribly useful and so for the time being the previous plot (Figure 2.22) will have to do.

The raw data for these failed and poor PCR rxns is here: 10_20_07_4volumeChecksReplicate251x3.

2.13 Making the bead/antibody surface using a small reaction volume

Ok, now back to where we were before I developed an obsession for trying to get enrichment using less beads and antibody rather than just using the huge amounts of antibody and bead that seemed optimal. I really want to be able to get optimal growth using a maximum of 50 μl of Dynal Beads. Initially, I thought lowering the amount of DNA was the key. But further tests suggested that this was only of limited help (Figure 2.19. However, it turned out that lowering the reaction volume looked like a more promising way to increase enrichment (Figure 2.22).

One really nice implication from the increase enrichment with smaller volume is that enrichment step might be really easy to adapt to a 96-well format. Towards that, I'm going to use a total volume

of 200 μl (the volume of a standard PCR strip and 96-well PCR plate) for this bead/antibody surface.

Thur Oct 18, 2007

I'll be using 29 μl of sheared sample 2 from sheared chromatin samples in section 2.12. So each enrichment will contain 29 μl sheared chromatin (2 μg), 71 μl of dilution buffer, and 100 μl of beads. Different amounts of beads will all be concentrated and resuspended in a final volume of 100 μl PBS + 0.5%BSA to maintain a consistent 100 μl of beads. The TE washes will be 200 μl .

I tested 12 total combinations. A 3x3 matrix of 1,2,5 μl antibody combined with 50, 100, and 200 μl of bead. In addition I added an extra 3 points in other spots of interest. The design was (excel file of design):

id	randId	antibody	bead
8	1	2	200
2	2	2	50
6	3	4	100
12	4	3	150
10	5	3	150
4	6	1	100
5	7	2	100
3	8	4	50
1	9	1	50
7	10	1	200
9	11	4	200
11	12	3	150

Sun Oct 21, 2007

I ran the two qPCR plates. Shortly before running the qPCR I realized I had only tiny amounts of primer left in my primer plate (lrpTechReplicate). Thankfully, I had a second aliquot in the freezer, which I thawed and fixed the problem I'd earlier had when I forgot to add the final dadA primer to its third technical replicate well in the primer plate. The two PCR plates were run back-to-back with 12 samples in each plate. Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see factorialBeadAntiSurface3.m for the script).

Brief Conclusions: When you compare the previous bead/antibody surface (Figure 2.24A) with the new low volume version (Figure 2.24B), it is quite clear that I achieved my goal of being able to get my original maximum enrichment using only 50 μl of antibody. However, what I wasn't expecting to find was that I could get even more enrichment by still using the high amounts of bead/antibody at the lower volume. The plots for dadA and serA get about 3x more enrichment at the low volume, and the enrichment is on a log scale so the old max enrichment for the two genes was around two-fold more than random, it is now around five-and-a-half fold more enriched than random. In these low volume plots, it appears that antibody is starting to play a more dominant role.

To keep costs down, I'm going to do the remaining optimizations using 3 μl of antibody and 100 μl of bead for each reaction. I think that will provide a nice balance between maximal enrichment and cost.

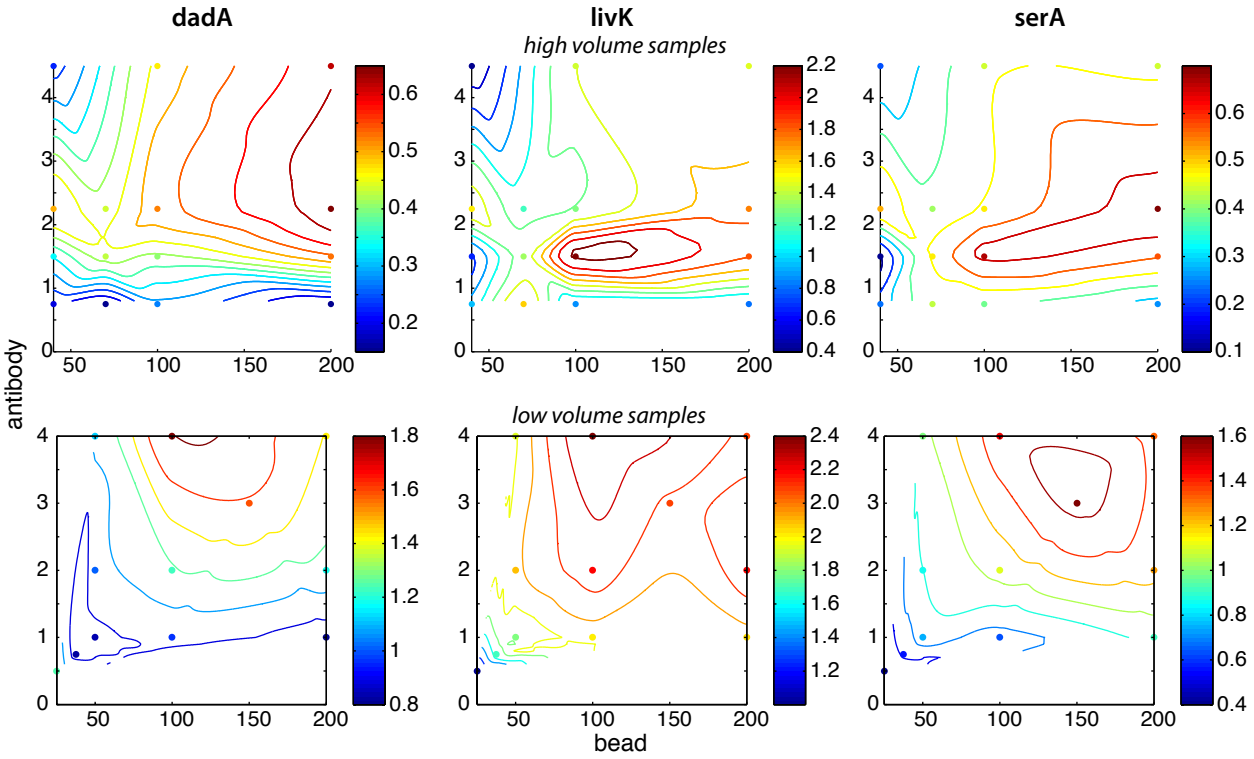


Figure 2.23: Bead/Antibody enrichment surface for high and low volume samples. The A) high volume samples panel is taken from Figure 2.18 but with the addition of the actual datapoints used to fit the contour.

2.14 Plate -vs- tube

Thur Oct 25, 2007

Now that the volume is down below 200 μl , we can attempt to enrich our targets using a plates or pcr-strips rather than with 1.5 ml tubes. Doing so would make running hundreds of samples much more tractable.

I purchased a 96-well Dynal magnet from Invitrogen (Dynal MPC-9600). Rather than just switching straight-away, I want to try the enrichment with tubes and with PCR-strips to see if there are any differences. For the tubes, I'm running the reaction just like before. For the PCR strips, I'm not rotating the samples. Rather I'm using the magnet to mix beads, as recommended in the MPC-9600 manual.

I used 32 μl of the sheared DNA (2 μg) from section 2.12, 100 μl of bead, and 58 μl of dilution buffer (for a total rxn volume of 190 μl). Samples 1-4 were the old tube way. Samples 5-8 were the new PCR-strip/96-well way. I changed buffers and eluted the samples in the PCR strips using a multichannel pipettor to verify that it isn't hard to work with the beads using a multichannel (I did the initial bead allocation with a single-channel though to prevent wasting beads).

Fri Oct 26 20:10:24 EDT 2007

I cleaned up the crosslink reversed samples with a Qiagen PCR purification kit. When I did the purification, I forgot to elute the samples and left them sitting in EB buffer for quite a while (10-20

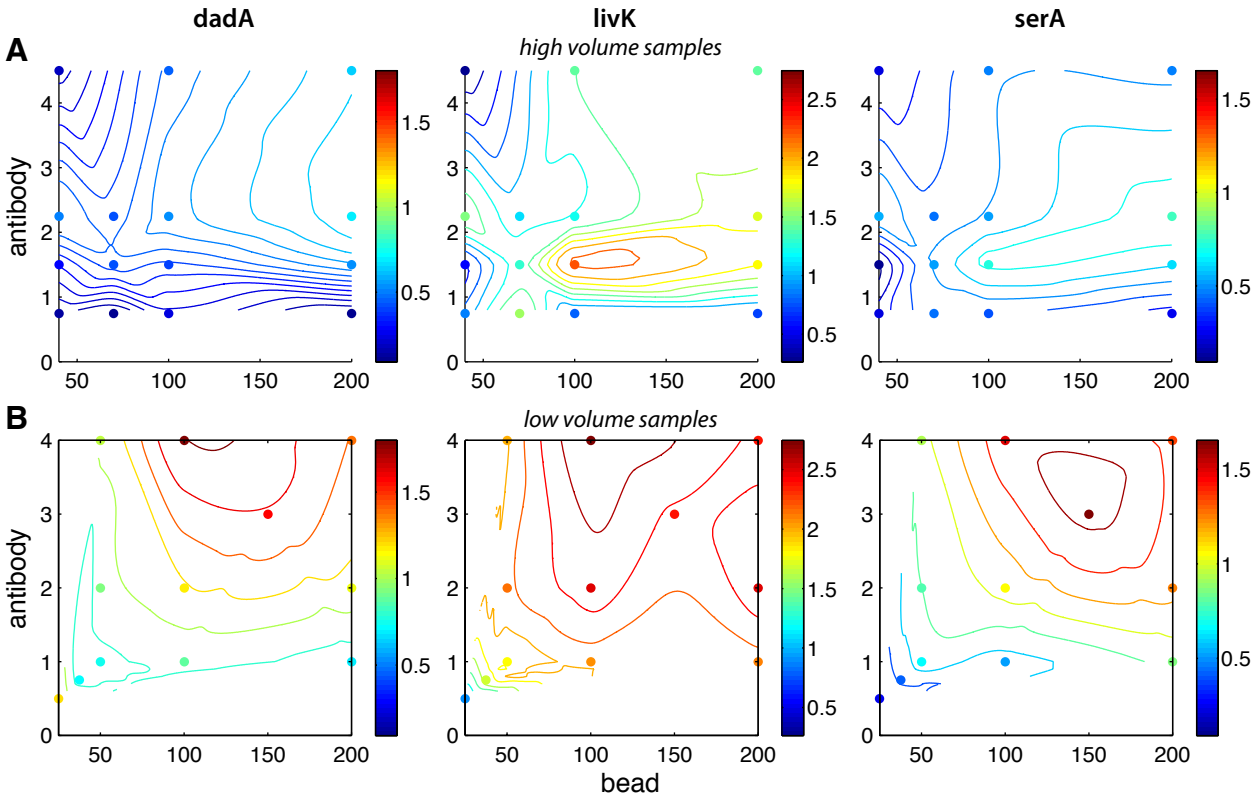


Figure 2.24: Bead/Antibody enrichment surface for high and low volume samples. The A) high volume samples panel is taken from Figure 2.18 but with the addition of the actual datapoints used to fit the contour. This plot is exactly the same data as Figure 2.23 except that both the high volume and the low volume are plotted using the same scale for the colorbar.

min); hopefully, that doesn't mess things up. I ran the qPCR on all 8 samples. Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see plate_vs_tube.m for the script).

The mean enrichment values for dadA, livK, serA, and pntA were 0.2760, 0.7373, 0.2934, and 0.1101 respectively. Although these were all positive, they were all much lower than the mean values from the bead/antibody surface in the previous section (1.1058, 2.0953, 0.9377, 0.5430). A two-tailed t-test between the plate and the tubes was not significant for any of the four tested genes pvals=(0.1681, 0.0935, 0.1490, and 0.4282), suggesting that there was not difference between the plate and the tubes (if you just look at the mean and ignor the noise, the plate did slightly better).

Brief Conclusions: It looks like there is not difference between the tube and the plate which is good – high-throughput here I come! However, the whole-experiment is tainted by the fact that the enrichment values for all four genes dropped by so much. I think the answer lies in the Ct values. The Ct values were *much* lower for this experiment than for all previous ChIP experiments I've run – ever. The mean Ct was 17 for this experiment whereas for the previous experiment is was 22. Since all four genes had positive enrichment, clearly I'm picking up signal, but it looks like the signal is masked by background noise. I'd guess I'm picking up DNA from somewhere? Next

time I run a ChIP experiment, I must make fresh dilution buffer, fresh elution buffer, fresh 0.5% BSA/PBS buffer, use a new box of qPCR master mix, clean my bench and pipettes throughly with DNA-away, and use a new source of DNA-free water.

How much this background signal interfered with my t-test is not clear. However, for now I'm going to move onward and assume moving to a plate doesn't really matter. The major difference between the two was that I mixed with rotation for the tubes. But with a volume of 200 μl the rotation really doesn't mix much anyways. Sunday I hope to make the formaldhehyde/shearing surface that I thought I'd make about a month ago...

2.15 Making the formaldehyde/shearing surface

Tues Oct 2, 2007

The last surface I want to make is the optimize formaldehyde concentration and shearing amount. My previous factorials strongly showed the influence of formaldehyde and hinted that there might be some interaction with shearing. Now I'm going to test this in more detail. I have two goals:

1. can we increase our ChIP enrichment still further by optimizing shearing and formaldehyde
2. shearing is still the least fun part of the protocol, so I'd like to the least amount of shearing that will give me an optimal result

I initially thought I'd use a 4×4 matrix of formaldhyde × shearing. However, I'm more concerned with formaldehyde, so I decided to do a 4x3 matrix. I added an addition 4 points at the far corners to try and help determine the boundards of the formaldehyde/shearing plot. The four formaldehyde values were: 0.5, 1, 2, 4%. The three shearing values were 1, 2, and 3x30secx20% power. The boundaries four points were a two-by-two matrix of formaldehyde=(0.1% and 8%) and shearing (1x30secx20% power and 6x30secx20% power).

The final set of 16 experiments was:

id	randID	formaldehyde	shearing	formaldhyde %	ml formaldehyde in 15ml total vol
5	1	0.5	3	0.1	0.041
12	2	4	4	0.5	0.2
1	3	0.5	2	1	0.405
7	4	2	3	2	0.81
3	5	2	2	4	1.62
4	6	4	2	8	3.24
15	7	0.1	6		
11	8	2	4		
8	9	4	3		
16	10	8	6		
2	11	1	2		
10	12	1	4		
9	13	0.5	4		
13	14	0.1	1		
14	15	8	1		
6	16	1	3		

Tues Oct 3, 2007

I cleaned (Qiagen PCR; elute 50 μl) and spec'd the samples and ran all 16 onto a 1% agarose gel for 40 min at 110V (Figure 2.25). The spec values are (raw spec data):

Sample ID	Date	Time	ng/ul	260/280	260/230
1	10/3/07	3:00 PM	147.56	1.89	2.01
2	10/3/07	3:01 PM	140.23	1.86	1.93
3	10/3/07	3:01 PM	131.66	1.86	1.91
4	10/3/07	3:02 PM	120.89	1.87	1.94
5	10/3/07	3:02 PM	119.08	1.87	2.06
6	10/3/07	3:05 PM	104.88	1.83	1.95
7	10/3/07	3:09 PM	112	1.86	1.96
8	10/3/07	3:10 PM	118.57	1.82	1.93
9	10/3/07	3:11 PM	105.18	1.84	1.93
10	10/3/07	3:12 PM	91.33	1.78	1.74
11	10/3/07	3:13 PM	100.53	1.85	2.03
12	10/3/07	3:14 PM	107.44	1.82	1.88
13	10/3/07	3:14 PM	108.62	1.82	1.93
14	10/3/07	3:15 PM	90.67	1.82	1.81
15	10/3/07	3:16 PM	31.02	1.78	1.73
16	10/3/07	3:16 PM	102.05	1.82	1.96
test1	10/3/07	3:19 PM	220.22	1.82	1.64
test2	10/3/07	3:20 PM	217.05	1.81	1.53
test3	10/3/07	3:20 PM	223.82	1.78	1.37

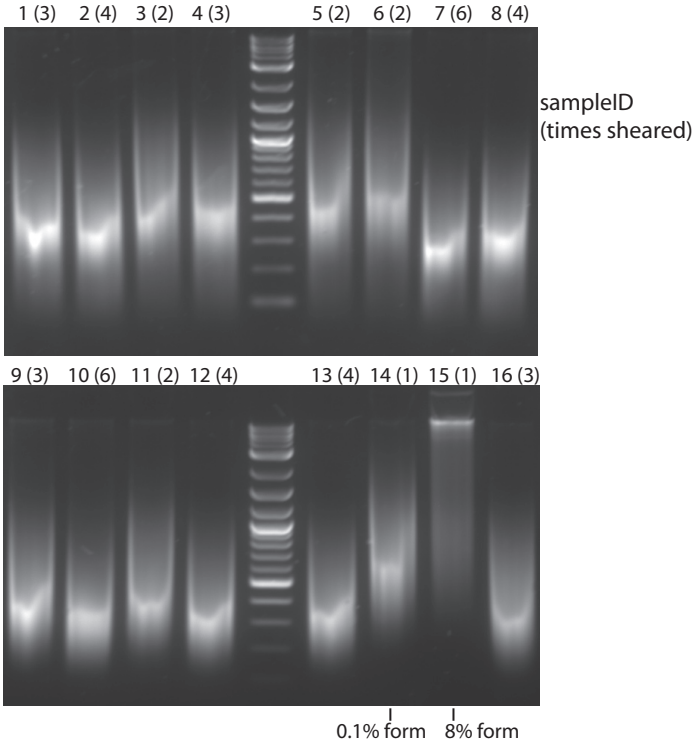


Figure 2.25: Sheared chromatin for different formaldehyde and bead concentrations

Brief Conclusions: The yields of about 100 ng/μl are about half of my normal yields. I'm not sure why, as the OD I grew the cells to was the same as normal. The three test spec readings above were run to make sure that the lower yields weren't resulting from errors with the plate reader (the tests seemed fine). Note random sample 15 had an extremely low yield. This was the high formaldehyde (8%) low shear sample.

Sun Oct 28, 2007

After a long break to sort out all kinds of stuff relating to the bead/antibody/DNA amount/and reaction volume, I'm finally read ready to enrich these guys. Using the small volume, 96-well approach developed above, I enriched each sample using 2.5 μl antibody and 100 μl of Dynal beads. For samples 1-14,16, I used the average of those sample DNA concentrations to estimate the amount of sheared chromatin to use for 2 μg (35 μl), the remaining 55 μl was dilution buffer. For the weak sample 15, I just used the maximum of 90 μl chromatin. When I washed, I resuspended the 100 μl of bead into 90 μl 0.5% BSA/PBS to reduce the reaction volume a bit. The total volume (not counting the antibody) for each of the samples was then 180 μl .

Antibody and bead incubations were both 40 minutes. And elution was for 15 min at 65C in a water bath in 180 μl of Dynal elution buffer. The samples were placed at 65C overnite to remove crosslinks.

Mon Oct 29, 2007

I added 4 μl proteinase K, incubated for 1 hr at 55C, cleaned up all 16 samples with Qiagen PCR purification columns, and eluted into 100 μl EB buffer.

I ran a qPCR on samples 1-8.

Tues Oct 30, 2007

I ran a qPCR on samples 9-16.

A plot of the results is shown in Figure 2.26.

Click for a little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see form_shear_original.m for the script; the two qPCR files are 10_29_07_shear_formaldehydePlate1.txt and 10_30_07_shear_formaldehydePlate2.txt).

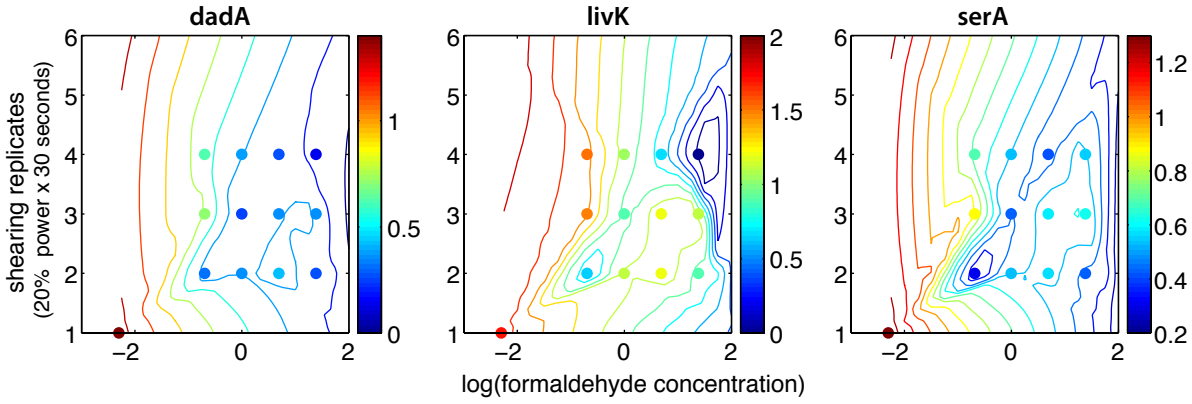


Figure 2.26: It seems that lower formaldehyde concentrations are enriching better.

Brief Conclusions: These results contradict my earlier formaldehyde result from the first factorial experiment where the 1% formaldehyde produced a boost of around 1-log more enrichment than the 0.1% formaldehyde (see section 2.2.9 on page 119). The lowest formaldehyde/shearing combination performed best. Kinda sucks because I feel stupid for checking all of those points all over the place when the bottom left point is the best! (Ideally the best point would be in the middle so I could see the performance of the parameter space around the optima).

However if this result proves true, it does have a couple nice implications: 1) I don't have to shear as much, which is good because shearing takes forever and is horribly boring; 2) I can use lower formaldehyde which opens the possibility that it'd be easier to cut the crosslinked DNA.

2.15.1 adding more low formaldehyde concentration data points

Because the results in the previous section contradict my earlier factorial result, I decide to try and fill in the form/shearing matrix with some additional low-concentration formaldehyde and low-shearing datapoints.

The randomized experimental design is (excel file of experimental design):

low-concentration formaldehyde shear surface design			
sampleId	**randomId**	**formaldehyde**	**shearing**
1	6	0.033	1
2	3	0.1	1
3	7	0.3	1
4	8	0.6	1
5	5	0.033	2
6	1	0.1	2
7	2	0.3	2
8	4	0.6	2

formaldehyde percent	**formaldehyde in 15 ml**
0.6	240 ul
0.3	120 ul
0.1	40 ul
0.0333	14 ul

Wed Oct 31, 2007

started overnite culture of LrpB

Thur Nov 1, 2007

I grew four cultures in LB in 50 ml baffled flasks from a 1:100 dilution of the overnite culture for 3hr 30 min to an background subtracted OD600 of: 0.469, 0.479, 0.474, 0.485 (0.01 μM IPTG was added after 1hr of growth). Two samples were taken from each of the four cultures to make the eight total samples in the experimental design above. Formaldehyde concentrations and shear where done according to the experimental design table above. 25 μl H_2O and 5 μl proteinase K was added to each 100 μl of each sample, prior to placing them at 65C overnite to reverse crosslinks. The remaining 900 μl of sheared chromatin was placed at -20C.

Fri Nov 2, 2007

I cleaned up the overnite crosslink-reversed samples using a Qiagen PCR purification kit, and ran them on a 1.5% agarose gel (Figure 2.27). Yields were:

overnite crosslink reverse with proteinase K			
Sample	**DNA (ng/ul)**	**260/280**	**260/230**
1	125.8	1.85	1.93
1	138.0	1.85	1.80
2	142.3	1.82	1.62
3	172.4	1.84	1.73
4	203.5	1.88	1.98
5	144.1	1.87	1.93
6	149.1	1.85	1.66
7	169.9	1.88	1.90
8	149.5	1.87	2.03

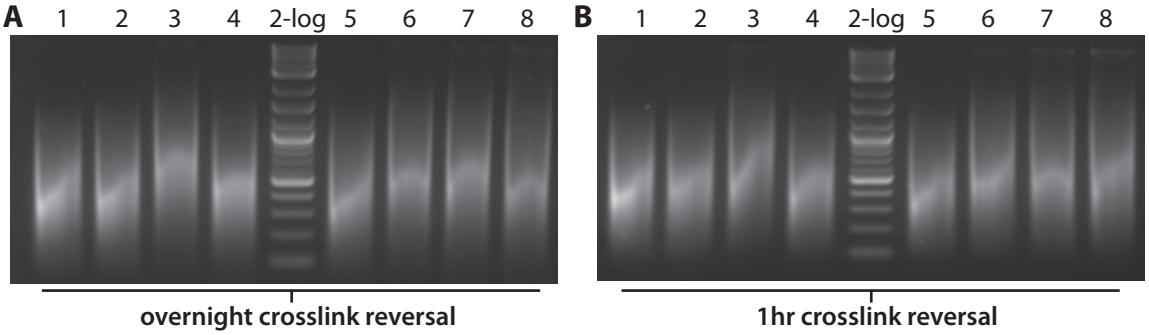


Figure 2.27: Crosslink for all 8 samples were reversed at 65C with proteinase K overnite (A) or for 1 hr (B).

I used the median DNA concentration value from the table above (149.3) to calculate the amount of chromatin to use for each IP (27 μl chromatin, 53 μl dilution buffer). I used 100 μl of beads washed 1x in 0.5% BSA/PBS and resuspended in 100 μl of 0.5% BSA/PBS.

I recorded the time necessary to prepare the beads in 96-well format: 15 minutes.

Sat Nov 3, 2007

I ran the qPCR for the 8 new formaldehyde/shearing concentrations and I combined the new results with the previous results (for 24 total formaldhyde/shearing combinations). A plot of the results is shown in Figure 2.28. Click for a little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see form_shear.m for the script; the newest plate is 11_03_07_shear_formaldehyde_plate3.txt and the previous qPCR files are 10_29_07_shear_formaldehydePlate1. txt and 10_30_07_shear_formaldehydePlate2.txt).

Brief Conclusions: With the addition of the additional 8 samples (Figure 2.28), we have further support that the lower-formaldehyde and shearing amounts are the best performing – contradicting the earlier factorial result.

2.15.2 Do I really need to reverse the crosslinks overnite to quantify the amount of sheared chromatin

Given that the 96-well format allows my to do the enrichment stuff in less than two hours, if I didn't have to reverse the crosslinks overnite to quantify the amount of sheared DNA, I could do

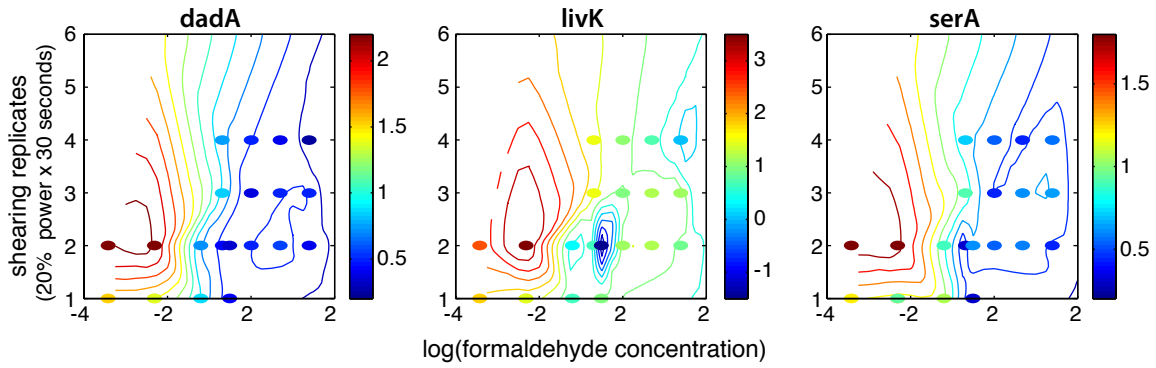


Figure 2.28: Further confirmation that the lower formaldehyde concentrations are performing best

the entire protocol in almost a day (at the end of which I was reverse the crosslinks on the *enriched* samples, and clean them up + qPCR them the following morning = 1.5 days).

To test this, I took the eight samples from the low-concentration formaldhyde experiment above and I took an extra 100 μl to quantify where I incubated them at 65C with the proteinase K for only an hour before I cleaned them up with a Qiagen PCR purification kit. I also took the normal 100 μl sample and reversed the crosslinks the typically overnight way with proteinase K. The yields for both methods are below (all samples were eluted into 50 μl of EB buffer):

Thur Nov 1, 2007

1 hr crosslink reverse with proteinase K			
Sample	**DNA (ng/ul)**	**260/280**	**260/230**
1	125.8	1.85	1.93
2	113.5	1.82	1.81
3	127.8	1.82	1.81
4	181.5	1.82	1.90
5	123.5	1.81	1.84
6	119.4	1.83	1.83
7	138.7	1.84	1.86
8	144.3	1.83	1.85

Fri Nov 2, 2007

overnite crosslink reverse with proteinase K			
Sample	**DNA (ng/ul)**	**260/280**	**260/230**
1	138.0	1.85	1.80
2	142.3	1.82	1.62
3	172.4	1.84	1.73
4	203.5	1.88	1.98
5	144.1	1.87	1.93
6	149.1	1.85	1.66
7	169.9	1.88	1.90
8	149.5	1.87	2.03

This the perfect kinda data for a paired t-test (I thought I'd live my whole life doing only unpaired t-tests). Comparing the 1hr with the overnite: pval = 7.8443e-04. With unpaired t-test pval = 0.0426; So it seems like there is a definite difference between overnight and 1 hr crosslink reversal

for DNA quantification. However, the means – 134 and 159 for 1 hr and overnite respectively – only differ by 15%.

Brief Conclusions: Given the decent robustness to changes in chromatin amount (e.g. 4-fold drop does affect things a bit, but 2-fold doesn't really change enrichment much), it might be possible to skip the DNA concentration entirely after a certain type of sample at a particular OD has already been run, you could just use the previous quantification values and things should be fine.

However for new samples or just to be safe, a 1 hr crosslink removal with proteinase K only differs from an overnite removal by 15%. 15% over or under is not going to make a noticable difference in the ChIP protocol, so I'm just going to switch over to using a 1 hr crosslink reversal for the DNA quantification step. This change will allow me to run the entire protocol (besides the final qPCR) in a day. It I really wanted to be closer to my target chromatin amount, I suppose it would be better to scale the 1 hr value by 1.15, but I'm not going to bother. Finally, the shearing range looks the same for the overnight and 1 hr crosslink reversals (Figure 2.27), so for checking the shearing range 1 hr looks fine as well.

2.16 testing the optimized protocol on a transcription factor besides Lrp

It's time for the real test: how well does our optimized protocol work on interactions it wasn't optimized for.

Sat Nov 3, 2007

I started overnite cultures of LrpB, PdhR, and FecI in LB.

Sun Nov 4, 2007

I started 9 cultures (3 of each) from 1:100 dilution of the overnites into 50 ml LB in a 250 mlbaffled flask. After one hour, I added 0.01 μM IPTG. I ran the latest version of the protocol using the single-day grow, lyse, shear, quantify, IP. This single-day protocol is possible because the IP is much faster in 96-well format and because the DNA quantification with 1 hr crosslink removal is almost the same as with overnite.

I timed *everything* to get an estimate of how long things would take.

The (excel file of experimental design for these 3 TFs with 6 replicates).

After 3 hr 10 minutes the samples were dense enough to shearing:

randCultureID	gene	OD600
1	fecI	0.489
2	pdhR	0.633
3	fecI	0.467
4	pdhR	0.638
5	fecI	0.476
6	lrp	0.628
7	pdhR	0.545
8	lrp	0.549
9	lrp	0.534

I took two 15 ml samples from each culture (18 total) into 15 ml centrifuge tubes. I used 40 μl of formaldehyde (0.1%), quenched with 750 μl glycine, washed 2x in PBS and sheared each sample with the Branson Digital Sonifier 250 for 2x20%x30sec.

I then did a 1 hr proteinase K (5 μl) crosslink reversal at 65C for one hour. I spec'd all 18 samples with the nanodrop:

Random Sample ID	Date	Time	ng/ul	260/280	260/230
1	11/4/07	6:14 PM	217.84	1.78	1.5
2	11/4/07	6:14 PM	202.72	1.67	0.95
3	11/4/07	6:15 PM	198.96	1.77	1.35
4	11/4/07	6:15 PM	202.71	1.79	1.39
5	11/4/07	6:16 PM	116.76	1.61	0.97
6	11/4/07	6:17 PM	243.05	1.72	1.16
7	11/4/07	6:17 PM	216.92	1.78	1.51
8	11/4/07	6:17 PM	202.43	1.75	1.34
9	11/4/07	6:18 PM	180.97	1.77	1.36
10	11/4/07	6:18 PM	222.51	1.81	1.6
11	11/4/07	6:19 PM	189.39	1.76	1.34
12	11/4/07	6:19 PM	196.88	1.78	1.38
13	11/4/07	6:20 PM	192.58	1.75	1.31
14	11/4/07	6:20 PM	180.55	1.77	1.33
15	11/4/07	6:20 PM	194.76	1.78	1.36
16	11/4/07	6:21 PM	176.63	1.8	1.55
17	11/4/07	6:21 PM	198.2	1.77	1.41
18	11/4/07	6:22 PM	152.04	1.79	1.65

The median values for fecI, pdhR, and lrp respectively were: 199.8, 196.5, and 195.8 ng/μl . I used 20 μl of each for the enrichment combined with 60 μl dilution buffer.

Timings:

Step	Labor time	Total time	Description
Growth	45 min	3 hr 10 min	
Crosslinking	1 hr 45 min	1 hr 45 min	
Shearing	1 hr 15 min	1 hr 15 min	
Quantification crosslink reverse	15 min	1 hr 15 min	
Quantification cleanup/spec	60 min	60 min	
Bead prep	8 min	8 min	
antibody incubation	10 min	40 min	
bead incubation	5 min	40 min	
2x TE wash	10 min	10 min	
final DNA cleanup protK	10 min	1 hr min	
final DNA cleanup Qiagen purificiation	1 hr	1 hr	
qPCR	45 min	2 hr	

total protocol time: 11hr 5 min on day one; 4hr day two

2.16.1 primer plates for the three transcription factors

Mon Nov 5, 2007

I need some primer plates to run the qPCR reactions. In order to compare with the original protocol and results that I generated with the PLoS paper, I'm going to use the original set of random genes not the newer set. I decided to use 14 primer pairs for each transcription factor.

PdhR primer plate							
-	1	2	3	4	5	6	7
A	gcl	mog	pinO	idnD	yhaF	nhaA	aceE
B	aimA	goaG	kdtB	yagG	citC	fruK	empty

Lrp primer plate							
-	1	2	3	4	5	6	7
A	gcl	mog	pinO	idnD	yhaF	nhaA	serA
B	aimA	goaG	kdtB	yagG	citC	fruK	livK

FecI primer plate							
-	1	2	3	4	5	6	7
A	gcl	mog	pinO	idnD	yhaF	nhaA	fecABCDE
B	aimA	goaG	kdtB	yagG	citC	fruK	fecIR

I used my typical concentration of 2 μM for each plate and made 400 μl of each primer pair in TE.

2.16.2 qPCR of targets of the three transcription factors

Mon Nov 5, 2007

I ran 16 samples on two qPCR plates. Unfortunately, I screwed up the ordering a bit and put the wrong primers in for a few samples which ruined them (I could run again later with correct primers).

In total I ran 5xFecI, 3xPdhR, and 5xLrp. Upon analyzing these samples, I decided to quit and not finish the remainder, as clearly, the results for PdhR and FecI (and even Lrp for that matter) were not as good as the original results for the unoptimized protocol. The serA and fecIR primers were also screwed up (I must have added them to the plate wrong). I also never bothered to run the sheared DNA on a gel to check shearing range.

Info to obtain the raw data for these make-me-unhappy results. Click for a little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see lrp_new_way.m fecI_new_way.m and pdhR_new_way.m for the scripts; the qPCR data is in the same directory files: fecI_new_way_qPCR.txt lrp_new_way_qPCR.txt pdhR_new_way_qPCR.txt

transcription factor	log(enrichment) old protocol	log(enrichment) new protocol
FecI	2.67	0.35
Lrp	3.13	2.2
PdhR	1.47	1.77

Brief Conclusions: Mon Nov 19 19:01:03 EST 2007

Clearly this was almost the worst possible result. This whole time I'd been optimizing with Lrp alone, I was always a little afraid my optimizations would be Lrp specific. But I assumed/hoped the process was pretty general and the optimizations would at least partially apply to other transcription factors (or at least not make them worse).

I went into a brief experimenter's depression after these results, which is why I'm adding these experiments to this notebook about 2 weeks after I did them. The thing about the results that most surprised me was how well I'd done with the original protocol. I do remember that I took a couple of decisions along the way that were purposely negative (e.g. use dynal beads instead of agarose; shorter incubations) in order to speed up the protocol. However, I figured that the subsequent response surface optimizations of bead/antibody formaldehye/shear would compensate

for those losses and in the end I'd have a protocol that is good or better than the original but with much less labor and much less time required per sample.

I have acheived the goal of higher throughput (throughput is now 10x the original given the 1.5 day protocol and the 96-well format). But at what cost? The one positive thing I could find about this negative result was that at least I knew which factors I had purposely chosen the less optimal state for in the factorial screens. I decided to rescreen those with *two* transcription factors to try and see if I can make the optimizations less transcription factor specific and to try and raise the overall enrichment for known targets relative to random targets.

2.17 a factorial with two TFs: removing the lrp specific optimizations

Given the sad result in the previous section where all of my optimizations have still left me with worse performance than my original protocol, I need to determine where I went wrong over the course of all of these experiments.

First let's state the accomplishments:

1. the protocol is *much* faster (as little as 1.5 days compared with the original 5-7 days)
2. the throughput is *much* higher (32 samples/day compared with the original 3.6 samples/day)

Seconds let's state the bad news:

1. somewhere along the way in my pursuit of a faster/leaner/higher throughput protocol, the performance took a hit

2.17.1 strategy to regain the lost performance

Along the way, I purposely chose a few factor states that *decreased* performance. I did this when the decrease in performance resulted in a major reduction in protocol time or ease. My hope was that the subsequent response surface optimizations would bring the protocol up to at least as good as the original. Alas, this doesn't seem to be the case.

When I began preparing this data for publication, I made a composite excel table of all the factors I'd tested and the effect size and p-value of each factor. Now, I'm going to select a few of the most important factors where I *chose the worst factor state*, and I'm going to screen them again in a factorial experiment with both Lrp AND FecI (so I'll really be running two factorial experiments). I'm purposely NOT including PdhR, so I have a final TF to use for cross-validation. The factorial summarizing table is available in excel format; I didn't include it in this text, because it was too wide to fit.

The factors I decided to test were: formaldehyde (0.1% — 1%), incubation time (overnite — 2 hr), bead type (dynal — agarose), high salt wash (yes — no). Using the knowledge gained from the shearing/formaldehyde surfaces, I'm used 2x20%x30sec for the 0.1% formaldehyde and 4x20%x30sec for the 1% formaldehyde. I also chose not to add in all of the salt washes; since I'd already shown the LiCl salt wash to not be important, I chose to *only* add back in the high salt

wash (500 mM NaCl, 20 mM Tris, 2 mM EDTA, 1% Triton X-100, 0.1% SDS). For the agarose work, I'll work in 1.5 ml tubes with 500 μl volume for the dilution and washes. For the dynal work, I'll stick with the 96-well PCR-strip based method using the 180 μl dilution and washes from my previous optimizations. Finally, for the agarose beads I'll use the original amount of 10 μg sheared chromatin (since I never optimized the chromatin amount with agarose beads), and for the dynal beads I'll stick with the 2 μg optimized sheared chromatin value.

I ran these 4 factorials in a 8-run fractional factorial design. The design was randomized for the two TFs. excel file of the design. Here's the design:

fractional factorial design for Lrp and FecI						
randID	**ID**	**gene**	**form**	**incubation time**	**bead type**	**high salt wash**
1	15	fecI	0.1	overnight:2hr	dynal	no
2	10	fecI	1	40min:40min	agarose	yes
3	4	lrp	1	overnight:2hr	agarose	no
4	16	fecI	1	overnight:2hr	dynal	yes
5	1	lrp	0.1	40min:40min	agarose	no
6	6	lrp	1	40min:40min	dynal	no
7	5	lrp	0.1	40min:40min	dynal	yes
8	13	fecI	0.1	40min:40min	dynal	yes
9	14	fecI	1	40min:40min	dynal	no
10	2	lrp	1	40min:40min	agarose	yes
11	8	lrp	1	overnight:2hr	dynal	yes
12	7	lrp	0.1	overnight:2hr	dynal	no
13	9	fecI	0.1	40min:40min	agarose	no
14	11	fecI	0.1	overnight:2hr	agarose	yes
15	12	fecI	1	overnight:2hr	agarose	no
16	3	lrp	0.1	overnight:2hr	agarose	yes

2.17.2 running the factorial experimental design

Tue Nov 6, 2007

started overnites of lrp, fecI, lrp

Wed Nov 7, 2007

grew 1:100 dilution from overnite of Lrp and FecI into two 250 ml flasks of each (4 flasks total).

After 3 hr the two lrp samples were 0.694 and 0.658 (OD600), but the fecI samples were only 0.112 and 0.119; once I saw this, I remembered (and verified by digging through this lab book) that I'd had this problem when working on the PLoS paper ChIP samples (see the first chapter *Chromatin Immunoprecipitation* of this lab notebook to find other examples of slow fecI growth). So I processed the lrp samples first. After reaching the resuspension of 2xPBS washed cells into dilution buffer, I processed the fecI samples (4 hr 30 min of growth with OD600 of 0.254 and 0.247).

I sheared all of the samples together, using 2x shearing for the 0.1% formaldehyde samples and 4x shearing for the 1% formaldehyde samples (as described in section 2.17.1 above).

After the shearing, I used quick DNA concentration method (5 μl proteinase K and 25 μl H_2O at 65C for 1hr followed by Qiagen PCR purification), yields for the four samples were (eluted into 50 μl EB):

Sample	DNA (ng/ul)	260/280	260/230
fecI 0.1% formaldehyde	125	1.85	1.70
fecI 1% formaldehyde	112	1.78	1.48
lrp 0.1% formaldehyde	136.3	1.79	1.74
lrp 1% formaldehyde	148.7	1.82	1.63

The mean yield was 131 ng/μl . I was a little surprised to see that the fecI yields were so close to the Lrp yields, since the Lrp OD600 was almost 2x the fecI value.

Also, I should mention that I only have 2 samples for each TF here, whereas my factorial experiment called for 8 samples. That's because I pulled two samples from each of the different formaldehyde amounts. I did this A) because I'm lazy and I could grow/shear half of the cultures and B) because this ensures that I start with the exact same sheared chromatin for the enrichment protocols, so there should be less noise when comparing the wash and bead factors on the same TF/formaldehyde combination.

Before going home, I started the overnite incubations with the dynal and agarose overnite samples with 2 μg and 10 μg of sheared chromatin respectively (as described in section 2.17.1 above). I used 2.75 μl (3.3 μg) of antibody for each antibody sample. I left the other (40min:40min) samples in the fridge.

Thur Nov 8, 2007

I started the 2 hr bead incubations and then shortly thereafter set up the 40 minute antibody incubation followed by the 40 minute bead incubation. So both groups finished at the same time.

2.17.3 new lrp and fecI primer plates

Fri Nov 9, 2007

I wanted to make some better PCR primer plates for this fractional factorial experiments. In particular, I wanted to limit the qPCR error by using a qPCR technical replicate. To do this I moved from 11 random genes to 9 random genes (I just removed the last 3 genes from the list) and used 3 technical replicates of one target for each tf (fecABCDE for fecI and livK for lrp).

By switching to 12 primer pairs per TF, I also made it possible to run all 16 samples (8 lrp and 8 fecI) in a single 384-well plate.

The plates are:

Lrp techRep small primer plate						
-	1	2	3	4	5	6
A	gcl	mog	pinO	idnD	yhaF	nhaA
B	aimA	goaG	kdtB	livK	livK	livK

FecI techRep small primer plate						
-	1	2	3	4	5	6
A	gcl	mog	pinO	idnD	yhaF	nhaA
B	aimA	goaG	kdtB	fecABCDE	fecABCDE	fecABCDE

I used my typical concentration of 2 μM for each plate and made 400 μl of each primer pair in TE.

2.17.4 lrp and fecI fractional factorial results

Fri Nov 9, 2007

I ran the qPCR plate using the two new primer plates described in the previous section.

Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see fecI_lrp_factorial.m for the script).

Below are the raw enrichment values for each of the tested factor combinations:

FecI and Lrp fractional factorial results					
form (%)	**incubation time**	**bead type**	**high salt wash**	**livK enrich**	**fecABCDE enrich**
0.1	40min:40min	agarose	no	1.0	1.5
1	40min:40min	agarose	yes	1.8	2.3
0.1	overnight:2hr	agarose	yes	2.0	1.9
1	overnight:2hr	agarose	no	1.0	1.1
0.1	40min:40min	dynal	yes	*2.8*	*2.7*
1	40min:40min	dynal	no	1.8	0.9
0.1	overnight:2hr	dynal	no	-1.3	-0.5
1	overnight:2hr	dynal	yes	*2.9*	*4.1*

Brief Conclusions: I haven't bothered to do a proper factorial analysis of these results, because the general idea just jumps right out of the table above (particularly after I added some *italics* to help them jump). The best performing methods are the same for BOTH TRANSCRIPTION FACTORS!!!!!!

The top two solutions are particularly interesting and thankfully they both use dynal beads (so the protocol can remain 96-well without coming up with a 96-well version for agarose beads). With a few extra experiments or perhaps just additional analysis of this data, I could tease apart what really mattered. However, I'm going to just use the top two solutions as-is, because they have nice properties.

The top performing combination is 1%:dynal:overnite:wash, which is a 2.5 day protocol.

The second best performing combination is 0.1%:dynal:40min:wash, which is a 1.5 day protocol.

One clear thing to take away, it that the high salt wash looks like it is certainly a crucial element to increase signal/noise (though I'm still curious why the LiCl based wash didn't do the job when I tested that by itself; my guess is that the 0.1% SDS in the high salt is really removing the background binding).

2.18 testing the 2.5 day, top-performing protocol, on Lrp, FecI, and Lrp

In the previous section, I was able to obtain maximal enrichment for FecI AND Lrp by using 1% formaldehyde, dynal beads, with an overnight antibody incubation, 2 hr bead incubation and a high salt wash. To determine if this was a fluke and to get an estimate of the error of this protocol, I'm going to try again (like I did in section 2.16) to run 6 replicates for each TF using this new protocol.

Sat Nov 10, 2007

I started an overnite culture of lrp, pdhR, and fecI

Sun Nov 11, 2007

I grew 3 x 50ml 1:100 dilution LB cultures in 250 ml baffled flasks for each transcription factor. Once again the fecI grew slower, so I processed the FecI samples after crosslinking and washing the

Lrp and PdhR samples. Lrp and PdhR were grown 3 hr 20 min (0D600 0.77 0.751 0.767 for Lrp samples 1,2,6 and 0.726 0.642 0.68 for PdhR samples 5,7,8). FecI samples were grown 4 hr 40 min (OD600 0.46 0.506 0.399 for FecI samples 3,4,9).

excel file of the randomized experiment setups below

Randomized Growth setup	
randID	**sample**
1	lrp
2	lrp
3	fecI
4	fecI
5	pdhR
6	lrp
7	pdhR
8	pdhR
9	fecI

FecI, Lrp, PdhR six ChIP replicates	
randID	**sample**
1	fecI
2	fecI
3	pdhR
4	lrp
5	fecI
6	pdhR
7	lrp
8	lrp
9	pdhR
10	fecI
11	pdhR
12	pdhR
13	lrp
14	lrp
15	fecI
16	fecI
17	lrp
18	pdhR

One important thing to note. Upto this point, I'd always washed the cells 2x with PBS to remove trace amounts of PBS and Glycine before sonication in dilution buffer. However, I didn't have enough PBS or enough time to make more, so I washed 1x in 8 ml of PBS (normally I use 2x10ml). To maximize the amount of media/formaldehyde/glycine that I removed, I placed the centrifuge tubes upside down on a stack of paper towels for a minute or so..

I crosslinked with 1% formaldehyde, sheared 4x20%x30secs. I quantified the sheared chromatin by adding 5 μl proteinase K and 25 μl H_2O to 100 μl of each sample and incubating at 65C followed by a Qiagen PCR purification. Average DNA amount was 97.87 ng/μl .

Raw nandrop data in excel file

Chromatin yields were (remember I used 100 μl EB here instead of the normal 50 μl ; it was by mistake, but I just have to remember to scale everything appropriately):

Sample ID	ng/ul	260/280	260/230
1	63.67	1.79	1.73
2	73.82	1.79	1.8
3	69.24	1.75	1.88
4	63.9	1.8	1.73
5	71.97	1.73	1.71
6	80.36	1.82	1.79
7	132.77	1.85	1.78
8	79.25	1.82	1.66
9	99.12	1.82	1.57
10	109.41	1.9	1.69
11	123.87	1.86	1.66
12	126.3	1.87	1.71
13	127.44	1.86	1.66
14	139.56	1.86	1.68
15	86.77	1.89	1.47
16	77.02	1.83	1.49
17	120.02	1.81	1.6
18	117.13	1.86	1.7

I eluted into 100 μl EB. 20.5 μl of each sample (2 μg) was mixed with 59.5 μl of dilution buffer and 2.75 μl (3.3 μg) of Anti-Xpress antibody. The PCR strips were rotated overnite at 4C.

Mon Nov 19, 2007 I ran the gel (Figure 2.29) after the qPCR and the analysis were done just to do a sanity check that the shearing lengths were fine (they were).

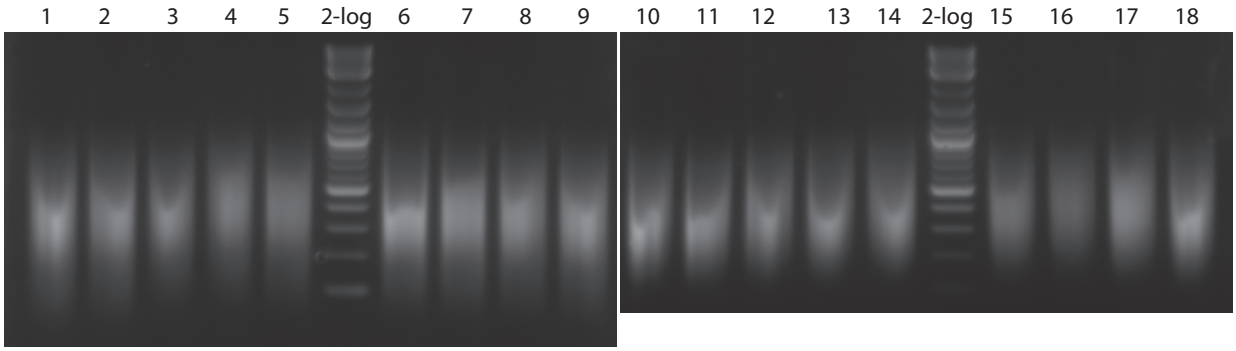


Figure 2.29: 1.5% agarose gel of sheared chromatin run for 37 minutes

Mon Nov 12, 2007

I added 100 μl of PBS/0.5%BSA washed beads and incubated for 2 hr at 4C. I washed the beads with 180 μl of high salt wash (5 min rotation at 4C). Finally I washed 2xTE at room temp and eluted into 180 μl of dynal elution buffer – leaving the samples at 65C for crosslink reversal.

2.18.1 more primer plates

Tues Nov 13, 2007

I'm going to make additional lrp, fecI, and pdhR primer plates to test these ChIP samples. I want to fit the PdhR and FecI samples on one plate, so I'm going to use the 11 random genes and one known target (aceE and fecABCDE for PdhR and FecI respectively). For lrp, I'm going to test most of the known targets I used to test my ChIP procedure for the PLoS 2007 paper in the first

chapter of this notebook. To test the technical/qPCR noise relative to the sample noise, I included 3 PCR technical replicates of pntA, serA, livK, gltB. I also included one replicate of stpA.

PdhR primer plate						
-	**1**	**2**	**3**	**4**	**5**	**6**
A	gcl	mog	pinO	idnD	yhaF	nhaA
B	aimA	goaG	kdtB	aceE	citC	fruK

FecI primer plate						
-	**1**	**2**	**3**	**4**	**5**	**6**
A	gcl	mog	pinO	idnD	yhaF	nhaA
B	aimA	goaG	kdtB	fecABCDE	citC	fruK

Lrp primer plate												
-	**1**	**2**	**3**	**4**	**5**	**6**						
A	gcl	mog	pinO	idnD	yhaF	nhaA	aimA	goaG	kdtB	pntA	citC	fruK
B	pntA	serA	livK	gltB	livK	serA	gltB	gltB	livK	pntA	pntA	serA

2.18.2 2.5 day protocol results

Tues Nov 13, 2007

I ran the 6 FecI and 6 PdhR samples in the 384-well qPCR plate together.

I ran the 6 Lrp reactions on a second plate.

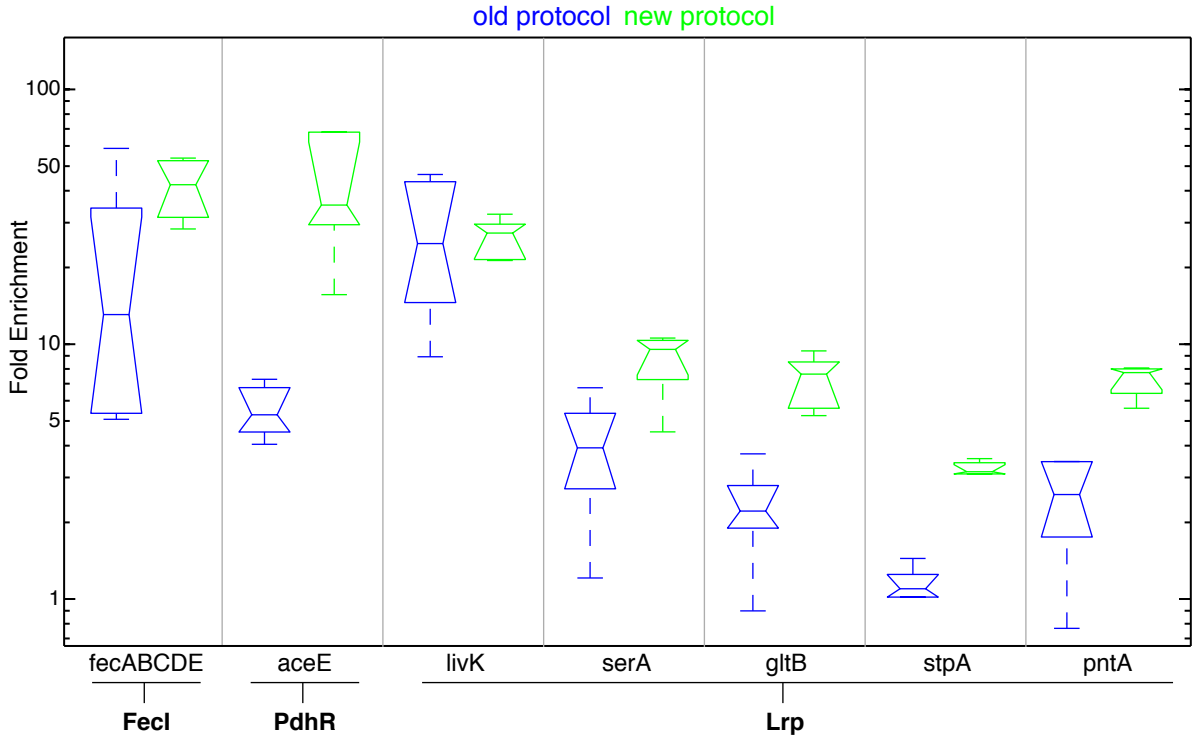


Figure 2.30: Boxplot of the old protocol vs the 2.5 day protocol.

Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see fecI_pdhR_new_way2.m and lrp_new_way_2ndTry.m for the scripts).

Brief Conclusions: It is done. The 2.5 day protocol after the final factorial optimization of FecI and Lrp is much faster, enrichs more, and is less noisy than the original protocol (Figure 2.30).

2.19 testing the 1.5 day, top-performing protocol, on Lrp, FecI, and Lrp

In section 2.17, I was able to obtain the 2nd best enrichment for FecI AND Lrp by using 0.1% formaldehyde, dynal beads, with a 40 min antibody incubation, 40 min bead incubation and a high salt wash. To determine if this was a fluke and to get an estimate of the error of this protocol, I'm going to run 4 replicates for each TF using this 1.5 day protocol.

Wed Nov 14, 2007

I started overnites of Lrp, FecI, and PdhR.

Thur Nov 15, 2007

I grew 50 ml cultures of all three TFs from dilutions of overnite in 250 ml baffled flasks. I used 1:100 dilutions for Lrp and PdhR. I used a 1:50 dilution for FecI (which grows slowly). I grew two of each TF. After 3 hr 10 min, the samples were taken with OD600: FecI (0.712, 0.673), Lrp (0.919, 0.896), and PdhR (0.825, 0.859) [I grew them a little longer than I would've liked].

I took two samples from each culture flask for 4 replicates total for each TF. The randomized order was:

lrp, lrp, fecI, pdhR, lrp, fecI, fecI, lrp, pdhR, pdhR, fecI, pdhR

I quantified the DNA by incubating 100 μl samples of sheared chromatin with 5 μl proteinase K and 25 μl H_2O for 1 hr at 65C followed by a Qiagen PCR purification. The DNA was quite concentrated (presumably because I let the cultures grow to a higher than normal OD600). I used 16 μl (appx 2 μg) with 64 μl dilution buffer for each immunoprecipitation. Based on the information from the nanodrop spec readings (raw spec data in excel):

Sample ID	ng/ul	260/280	260/230
1	241.19	1.83	1.81
2	185.94	1.83	1.75
3	271.84	1.8	1.56
4	247.77	1.83	1.83
5	281.99	1.84	1.88
6	268.81	1.81	1.67
7	279.45	1.8	1.57
8	280.24	1.83	1.83
9	237.46	1.83	1.86
10	232.56	1.84	1.85
11	252.77	1.84	1.79
12	225.22	1.84	1.83

Fri Nov 16, 2007

I cleaned up the samples and ran the qPCRs all on one plate using the same primer plates as I used in the previous section.

Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see old_new_compare_combine.m for the script). I analyzed the 1.5 day protocol and combined it with the original and the 2.5 day protocol (Figure 2.32).

Fri Nov 19, 2007

I ran the gel (Figure 2.31) after the qPCR and the analysis were done just to do a sanity check that the shearing lengths were fine (they were).

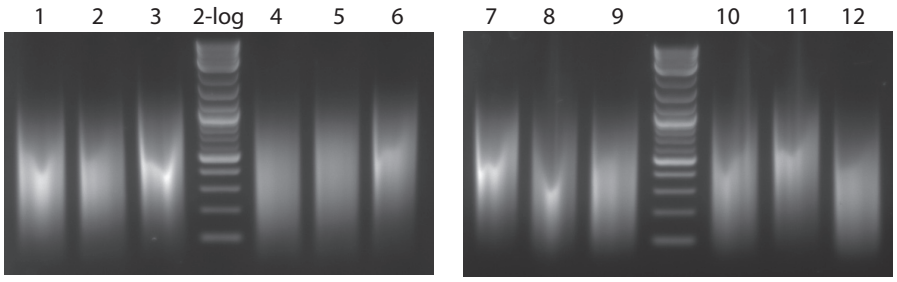


Figure 2.31: 1.5% agarose gel of sheared chromatin run for 30 minutes

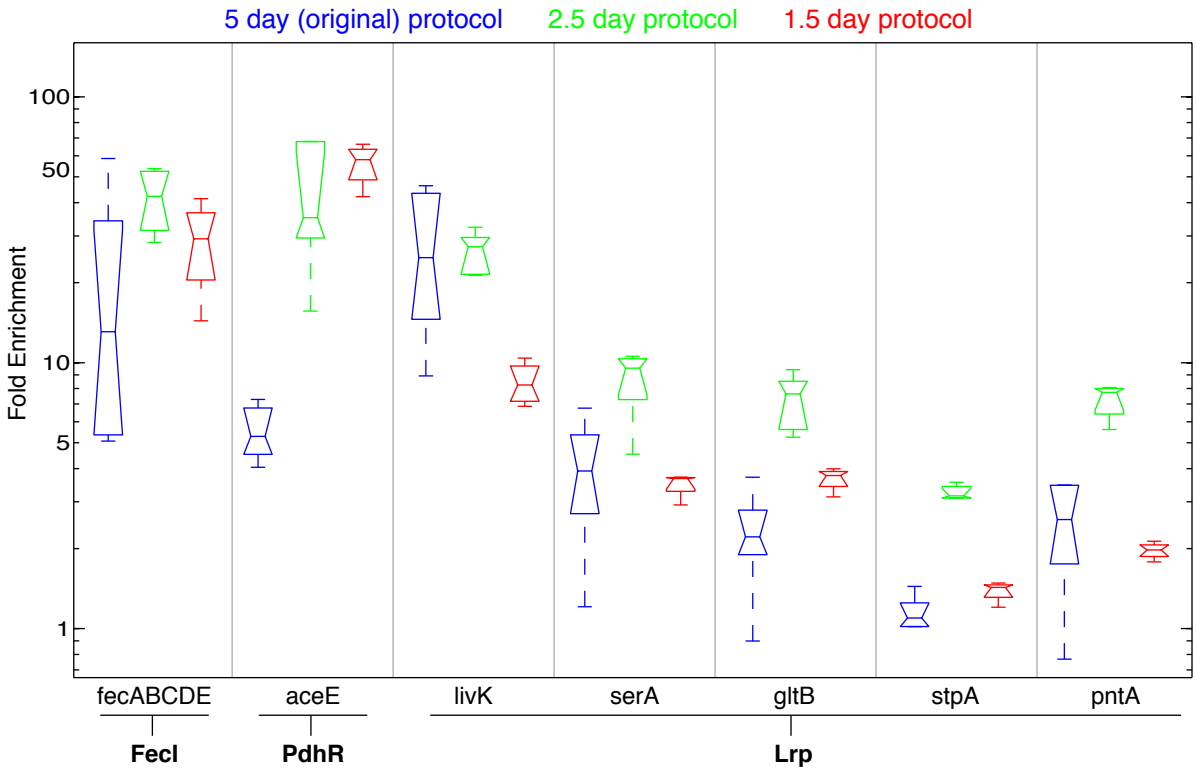


Figure 2.32: Boxplot of the old protocol vs the 2.5 day vs the 1.5 day protocol.

Brief Conclusions: The 1.5 day and 2.5 day protocols reflect the results obtained from the previous fractional factorial of Lrp and FecI quite well and both outperform the original protocol in general (Figure 2.32).

2.20 combinatorial epitope tags to improve consistency?

The optimizations thus far have exceeded my expectations. The factorial and response surface optimizations certainly helped move towards a much faster and better performing ChIP protocol. However one thing that I haven't address thus far is why some tagged transcriptions fail. The new protocols take the ChIP reactions that worked before and make them better, but how can we get the transcription factors that don't enrich with ChIP to work? In the past I've found on multiple attempts that LexA doesn't with ChIP (using both the original protocol and a partially optimized protocol about half-way through these factorial and RSM optimizations). I also have a tiny amount of effidence that PhoP and CysB don't work either (I tried them only one time). To determine the binding sites of all TFs, we're either going to need a gigantic custom built monoclonal library (my dream) we're going to need to figure out alternative for these tagged-TFs that don't enrich with ChIP.

My first guess as to why some TFs work and some TFs don't is that the tag is either inhibiting the binding of the TF to the genome or the TF protein folding is inhibiting the access of the antibody to the epitope tag. To test this idea, Ilaria (cloning master) Mogno, built four different combinations of epitope tags for three transcription factors (Figure 2.33). The clones are on an extremely low-copy plasmid (3-5 copies/cell) with kan resistance.

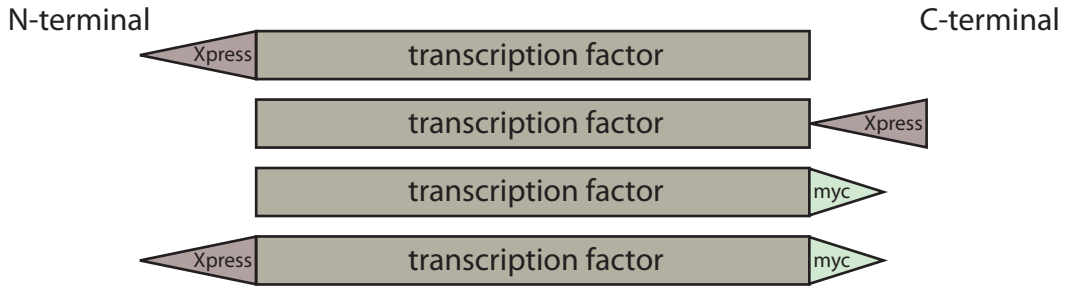


Figure 2.33: Ilaria made tagged versions of lexA, fecI, and lrp using four different locations and combinations of the Xpress tag and the myc tag

2.20.1 testing the four epitope tag combinations with lexA, fecI, and lrp

Sun Jan 27, 2008

I started 3 ml of overnite for each of the 12 strains.

Mon Jan 28, 2008

I began the 1.5 day ChIP protocol with all 12 samples at around 11:30AM. I used a 1:100 dilution for the lexA and lrp strains and a 1:50 dilution for the fecI strains. After 1hr 30min, I added 10 μM of IPTG (this was a half-an-hour later than I meant to add it). After 2hr 35 minutes, the cells had reached the appropriate OD600 of around 0.5-0.8. The OD600 (not background subtracted) and sample order is in the table below.

Experimental design for first epitope tag tests				
sampleID	**randomID**	**sampleName**	**OD 2hr 10min**	**OD 2hr 35 min**
7	1	XlrpN	0.439	0.62
8	2	XlexAN	0.425	0.617
2	3	XlexA	0.421	0.614
9	4	XfecIN	0.555	0.841
6	5	fecIN	-	0.825
1	6	Xlrp	-	0.638
4	7	lrpN	-	0.641
5	8	lexAN	-	0.544
10	9	lrpX	-	0.648
11	10	lexAX	-	0.62
3	11	XfecI	-	0.839
12	12	fecIX	-	0.851

For the strains with both an Xpress and a myc epitope, I used 2.75 μl (appx 3.3 μg) of *each* antibody (so twice as much total antibody).

Brief Conclusions: With the low copy plasmid, it appears that the fecI grows just as fast as the other two strains, so I should probably use 1:100 for all of the strains from now on.

Tues Jan 29, 2008

I cleaned up all of the 12 samples (24 total including the -antibody and +antibody).

I built a new primer plate that is essential the previous lrp primer plate with an additional row (C) of lexA targets in triplicate (for PCR technical replicates).

Lrp and LexA tag-test primer plate												
-	**1**	**2**	**3**	**4**	**5**	**6**						
A	gcl	mog	pinO	idnD	yhaF	nhaA	aimA	goaG	kdtB	pntA	citC	fruK
B	pntA	serA	livK	gltB	livK	serA	gltB	gltB	livK	pntA	pntA	serA
C	sulA	umuC	dinF	recA	sulA	umuC	dinF	recA	sulA	umuC	dinF	recA

Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see lexA_lrp_factorial.m for the script).

Brief Conclusions: Those preliminary qPCR results, hint that the lexA might be enriching for the first time! However, overall the qPCR reaction itself was pretty crappy (I was in a hurry and I don't think I filled the plate perfectly; quite a number of failed reactions).

repeating the qPCR reaction

Mon Feb 4, 2008

Given that so many of the PCR reactions failed in the previous qPCR plate, I decided to repeat the lexA lrp qPCR using the remaining ChIP DNA. I ran the exact same reaction, being extra careful no to make any mistakes while filling the plate.

Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see lexA_lrp_factorial_rep2.m for the script).

Brief Conclusions: qPCR reactions were much cleaner, and produced the same basic conclusions that the lexA might be working (and that the lrp works with the new tags as well). I'll add some figures and things after I get some replicates.

qPCR with fecI and the new tags

Tues Feb 5, 2008

I ran a qPCR of the fecI samples using the following fecI primer plate:

FecI tag-test primer plate							
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**
A	gcl	mog	pinO	idnD	yhaF	nhaA	*fecABCDE*
B	aimA	goaG	kdtB	*fecABCDE*	citC	fruK	*fecABCDE*

Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see fecI_IM_tags.m for the script).

Brief Conclusions: More positive results, seems the new tag library works work fecI as well. I'm not sure about the tagged transcription factors that contain only an Xpress epitope, but both of the tagged TFs that contain a myc tag worked (the tag with *only* a myc tag rather than both a myc and an Xpress tag worked best for this replicate). The version with a single myc on the N-terminal worked best for both lexA and fecI; the lexA version with N-terminal myc didn't seem to work and the values looked pretty odd, so we'll have to see how this looks in another replicate. These results are certainly promising enough to merit another replicate.

adding a second replicate of the four epitope tag combinations with lexA, fecI, and lrp

Fri Feb 8, 2008

I ran the 1.5 day ChIP protocol version 1.1 with the growing a 1:100 dilution of the same 12 strains in LB using the following randomized order:

sampleID	randomID	sampleName	OD600
8	1	XlexAN	0.661
11	2	lexAX	0.612
9	3	XfecIN	0.613
12	4	fecIX	0.598
3	5	XfecI	0.633
1	6	Xlrp	0.675
6	7	fecIN	0.597
2	8	XlexA	0.555
10	9	lrpX	0.603
5	10	lexAN	0.577
7	11	XlrpN	0.605
4	12	lrpN	0.56

After 1 hr of growth 20 μM of IPTG was added to each sample (I meant to add 10 μM like last time, but I messed up and added 20 μM). Samples were taken for crosslinking at 2 hr 30 min at an OD600 of around 0.6 (see table above; values not background subtracted).

Average sheared chromatin was around 100 ng/μl , so I used 20 μl of sheared chromatin and 60 μl of dilution buffer for the immunoprecipitations.

Sat Feb 9, 2008

I cleaned up the 24 rxs and ran the qPCR plate for the lrp and lexA strains.

Sun Feb 10, 2008

I ran the qPCR plate for the fecI strains

Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see lexA_lrp_factorial_completeRep2.m and fecI_IM_tags_completeRep2.m for the scripts).

combining the combinatorial epitope tag results

I'm going to combine the two factorial results into a single table with the *median* value for each gene for each replicate. For replicate 1, I'm going to use the second qPCR rxn for lexA and lrp where far fewer rxns failed.

transcription factor	target	log enrich rep1	log enrich rep2
XlexA	sulA	-0.036	0.023
lexAX	sulA	0.027	0.014
lexAN	sulA	0.1801	0.985
XlexAN	sulA	0.328	0.533
XlexA	umuC	0.032	0.059
lexAX	umuC	-0.056	0.008
lexAN	umuC	-0.355	0.961
XlexAN	umuC	0.100	-0.101
XlexA	dinF	0.108	-0.071
lexAX	dinF	-0.032	0.084
lexAN	dinF	1.5172	0.879
XlexAN	dinF	0.867	0.650
XlexA	recA	-0.012	-0.086
lexAX	recA	0.006	-0.014
lexAN	recA	1.570	0.497
XlexAN	recA	0.143	0.343
fecIX	fecABCDE	0.080	0.011
XfecI	fecABCDE	0.024	0.078
fecIN	fecABCDE	1.722	2.659
XfecIN	fecABCDE	0.639	1.599
Xlrp	pntA	-0.085	-0.022
lrpX	pntA	0.086	-0.232
lrpN	pntA	0.362	0.105
XlrpN	pntA	0.168	0.120
Xlrp	serA	0.420	-0.083
lrpX	serA	0.462	0.233
lrpN	serA	-1.989	0.324
XlrpN	serA	1.032	1.060
Xlrp	livK	-0.742	0.046
lrpX	livK	0.789	0.343
lrpN	livK	-2.001	1.287
XlrpN	livK	2.485	2.544
Xlrp	gltB	1.001	-0.036
lrpX	gltB	0.301	0.044
lrpN	gltB	-4.666	1.000
XlrpN	gltB	0.922	1.265

I think the table makes more sense if sorted by tag:

transcription factor	target	log enrich rep1	log enrich rep2
XlexA	sulA	-0.036	0.023
XlexA	umuC	0.032	0.059
XlexA	dinF	0.108	-0.071
XlexA	recA	-0.012	-0.086
XfecI	fecABCDE	0.024	0.078
Xlrp	pntA	-0.085	-0.022
Xlrp	serA	0.420	-0.083
Xlrp	livK	-0.742	0.046
Xlrp	gltB	1.001	-0.036
lexAX	sulA	0.027	0.014
lexAX	umuC	-0.056	0.008
lexAX	dinF	-0.032	0.084
lexAX	recA	0.006	-0.014
fecIX	fecABCDE	0.080	0.011
lrpX	pntA	0.086	-0.232
lrpX	serA	0.462	0.233
lrpX	livK	0.789	0.343
lrpX	gltB	0.301	0.044
lexAN	sulA	0.1801	0.985
lexAN	umuC	-0.355	0.961
lexAN	dinF	1.5172	0.879
lexAN	recA	1.570	0.497
fecIN	fecABCDE	1.722	2.659
lrpN	pntA	0.362	0.105
lrpN	serA	-1.989	0.324
lrpN	livK	-2.001	1.287
lrpN	gltB	-4.666	1.000
XlexAN	sulA	0.328	0.533
XlexAN	umuC	0.100	-0.101
XlexAN	dinF	0.867	0.650
XlexAN	recA	0.143	0.343
XfecIN	fecABCDE	0.639	1.599
XlrpN	pntA	0.168	0.120
XlrpN	serA	1.032	1.060
XlrpN	livK	2.485	2.544
XlrpN	gltB	0.922	1.265

Brief Conclusions: It's pretty noisy, but clearly the XtfN and tfN work the best. I think I want to continue on with both versions and run some replicates using the 2.5 day protocol. With the current data is seems that fecI definitely works better with a N-terminal myc only; lrp works better with the the C-terminal Xpress and the N-terminal myc; and lexA I can't tell. In general it looks like the second rep worked better than the first perhaps because of the added IPTG? Perhaps I should beef it up even further with the 2.5 day samples (perhaps 50 μM or 100 μM).

2.20.2 testing the two best epitope tag combinations with the 2.5 day protocol and more replicates

Sat Feb 16, 2008

The XtfN and tfN designs seemed to work the best with the 1.5 day protocol (the others didn't work at all?). With the 1.5 day protocol, I only ran 2 replicates. This time I want to try these two best epitope combinations with the 2.5 day protocol and 4 replicates to compare the new tags with the old C-terminal Xpress tag in the pTrcHis vector [Invitrogen].

I grew up 2 replicates of each of the three TFs and each of the two epitope combinations (12 samples total) in 35 ml LB + kan (35 μg /ml) from a 1:100 dilution of overnite culture. I split each sample in two after they reached an OD600 of 0.5-0.7.

sample	**OD600 sample 1**	**OD600 sample 2**
lrpN	0.587	0.556
XlrpN	0.575	0.539
lexN	0.578	0.578
XlexN	0.544	0.583
fecN	0.562	0.577
XfecN	0.589	0.573

The 24 split samples were run in randomized order:

experimental design fo tfN and XtfN tags with 2.5 day ChIP		
sampleID	**randomID**	**sampleName**
17	1	fecN
22	2	XlexN
11	3	fecN
24	4	XfecN
16	5	XlexN
23	6	fecN
1	7	lrpN
20	8	XlrpN
6	9	XfecN
14	10	XlrpN
3	11	lexN
2	12	XlrpN
8	13	XlrpN
18	14	XfecN
5	15	fecN
4	16	XlexN
19	17	lrpN
13	18	lrpN
21	19	lexN
7	20	lrpN
9	21	lexN
12	22	XfecN
15	23	lexN
10	24	XlexN

Something a little different this time with the initial quantification, rather than cleaning up and quantifying all 24 samples (which would be excessive and costly), I just quantified the first 6 samples. I did so, and 17.3 μl of sheared chromatin corresponded to 2 μg . For each sample, I added 62.6 μl of dilution buffer per sample.

Running all 24 samples took 8 hr and 51 minutes on day 1. One minor mistake, I added Xpress antibody in the no-antibody well of sample 12 (which is a XtfN sample).

Sun Feb 17, 2008

Day two of 2.5 day ChIP: add beads, high salt wash, 2xTE wash, elution; no problems.

Mon Feb 18, 2008

Day three of 2.5 day ChIP; cleaned up all 48 samples (24 no-antibody and 24 antibody) with Qiagen PCR purification columns. I ran the first lrp and lexA plates with samples (7, 8, 10, 12) and (2, 5, 11, 16). I also all eight of the fecI samples (1, 3, 4, 6, 9, 14, 15, 22).

Tues Feb 19, 2008

Ran the final lrp and lexA samples (13, 17, 18, 20) and (19, 21, 23, 24).

Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see twoPointFiveDay_compare_combine_with_xpress.m for the script).

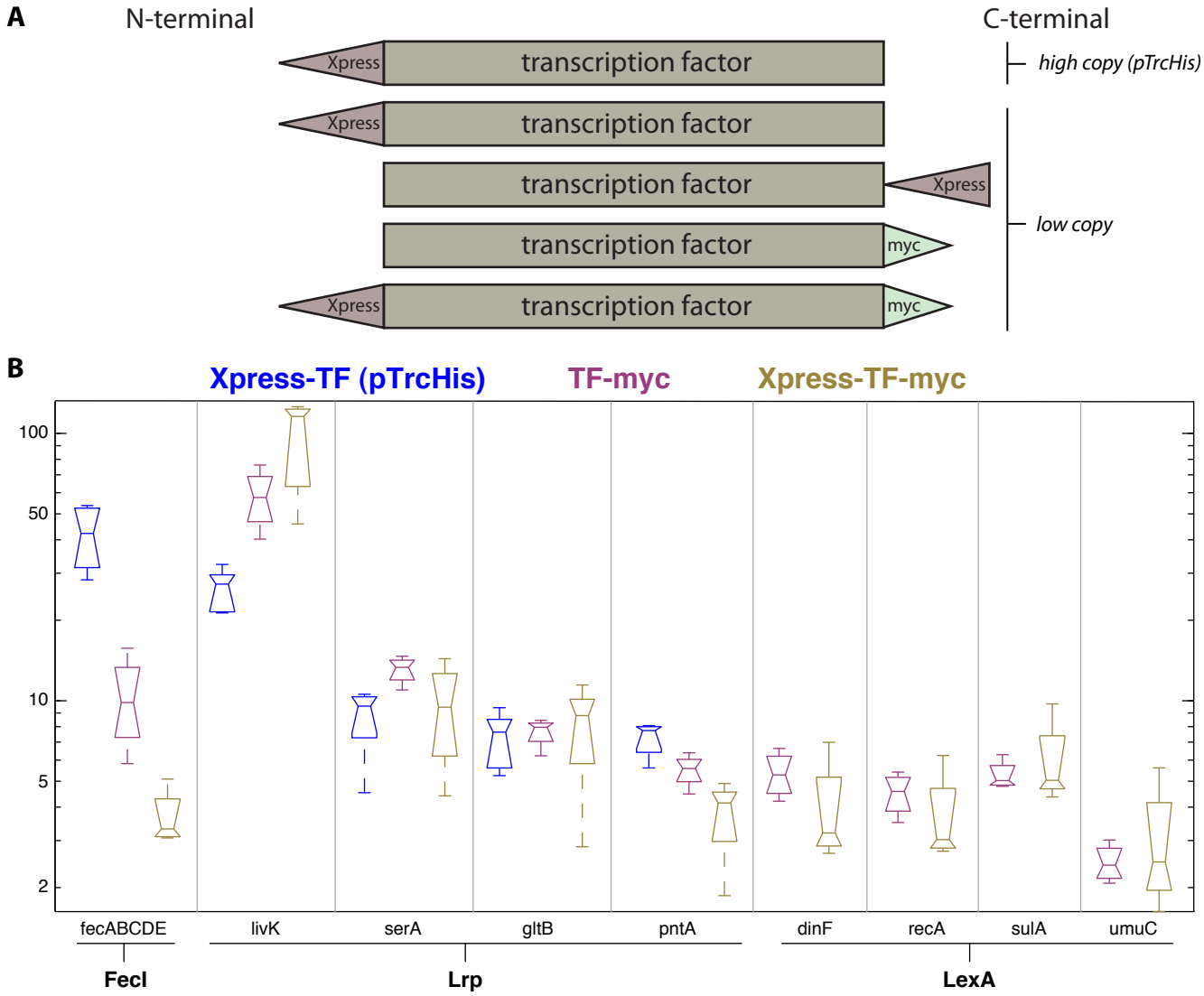


Figure 2.34:

Brief Conclusions: The higher replicate samples with the 2-day protocol show that the TF-myc and the Xpress-TF-myc variants work pretty well with the original Lrp and FecI targets in addition to work with the LexA targets which would enrich with the original vector (Figure 2.34).

2.20.3 expanding the new vector to new TFs

Fri Feb 22, 2008

Ilaria cloned cysB, oxyR, phoP, purR, and soxS into the TF-myc and Xpress-TF-myc. As a first pass I'm going to try the cysB, phoP, purR, and soxS strains. I've already tested cysB, soxS and phoP with the original high-copy pTrcHis vector and found they did not enrich known targets, so this will be an additional test to see if the new vector generalizes to a wider range of TFs.

I ran 4 replicates of each. I started 2 culture replicates of each TF and few them 2 hr and 30 min from a 1:100 dilution of overnite in LB. I split each sample in two before adding the formaldehyde.

sample	OD600 sample 1	OD600 sample 2
cysB	0.560	0.565
phoP	0.595	0.607
purR	0.538	0.574
soxS	0.515	0.518

The 16 split samples in randomized order:

experimental design fo tfN and XtfN tags with 2.5 day ChIP		
sampleID	**randomID**	**sampleName**
10	1	purR
12	2	purR
9	3	purR
6	4	phoP
7	5	phoP
4	6	cysB
13	7	soxS
3	8	cysB
11	9	purR
2	10	cysB
5	11	phoP
8	12	phoP
15	13	soxS
1	14	cysB
16	15	soxS
14	16	soxS

I used 2 μg of sheared chromatin 20.7 μl with 59.3 μl of dilution buffer. I used the 2.5 day protocol version 1.1 with 50 μM IPTG for TF induction.

Sun Feb 24, 2008

I ran all 3 qPCR plates. Plate 1 was cysB (A-H) and phoP (I-P); Plate 2 was purR (A-H) and soxS (I-P).

Click for little info about how I analyzed the above data and the exact matlab batch scripts I used are in the same directory (see purR_soxS_tfN.m cysB_phoP_tfN.m for the scripts)

Brief Conclusions: The cysB and purR worked. phoP and soxS did not. Overall the performance for the new vectors compared with the original is shown in the table below.

transcription factor	pTrcHis	TF-myc	targets enriched
fecI	Y	Y	fecABCDE
lrp	Y	Y	livK, serA, gltB, stpA, pntA
cysB	N	Y	cysB, cbl, cysK, tauA, cysP, cysD, cysJ
lexA	N	Y	dinF, recA, sulA, umuC
phoP	N	N	
soxS	N	N	
purR	-	Y	purE, purR, purM
pdhR	Y	-	aceE
% success	42.9	71.4	

2.21 END OF COMPLETED SECTIONS

Note that all sections below this one are ideas I've either not completed or I've decided not to complete.

2.22 Cool antibody factorial experiment to maybe try someday

2.23 Cloning lrpB into a myc-tagged plasmid

Thur Jun 21, 2007

I want to try lrp with the tag on the other end. I'm also going to use the myc tag instead of the Xpress tag. I ran the following PCR reaction for cloning in the pTrcHis from invitrogen: Easy-A master mix 10 μl , 7.5 μl H_2O , 2 μl primers (from 10 μM stock), 0.5 μl lrpB plasmid. Note that the lrp-myc primers remove the last codon of the gene so that the myc tag can be translated.

I used 1 μl of the PCR in the TOPO rxn, and I cloned the plasmid according to Invitrogen's instructions.

2.23.1 miniprepping and sequencing the new lrp plasmids

I grew up 3 lrp low-copy and 4 lrp-myc plasmids and miniprepped them. I eluted into 50 μl Yields were (lc = low-copy, m = myc):

Sample	DNA (ng/ul)	260/280	260/230
lc-A	69.2	2.16	2.19
lc-B	59.9	2.02	2.57
lc-C	93.5	2.11	2.36
m-A	150.3	2.02	2.26
m-B	100.2	2.05	2.21
m-C	111.7	2.00	2.27
m-D	104.5	2.02	2.11

I cut 10 μl of the miniprepped DNA to check the insert size (Figure 2.35). All of the insert sizes looked appropriate.

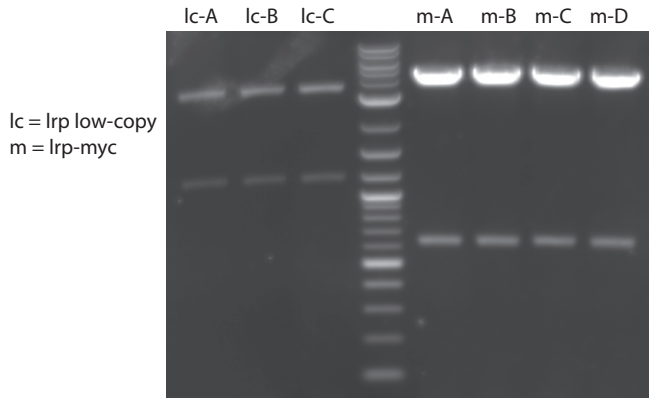


Figure 2.35: gel of cut lrp lowcopy and cut lrp myc

2.24 NOTES

matrix to test formaldehyde vs shearing:

form: 0.1, 1, 2, 4

shear: 2, 3, 4, 5

2.25 list of annoying things I'd like to check with a final factorial

1. do I really need to wash the beads with BSA? Can I just use PBS? Or citrine?
2. do I really need the two TE washes? or can I just add elution buffer and be done?
3. do a incubation time experiment where I check 5 min, 10 min, 40 min, and 80 min for a final sanity check if incubation time matters
4. can I do the bead/antibody incubation at RT?
5. do I really need to wash the cells 2x with ice cold PBS after formaldehyde and quenching? is one time enough? is the wash necessary at all? why not just add dilution buffer?
6. do I really need to place the dilution buffered beads at 65C (annoying with PCR strips)?
7. is proteinase K necessary during the crosslink reversals? (try both with 100 μl of sample; run a paired-ttest with the samples)
8. try the low pH buffer recommended by dynal
9. try the His Antibody + Xpress?

try: BSA beads (y/n), low pH beads (y/n), TE (2x/0x), wash with PBS (2x/1x), proteinase K (y/n), low pH dilution (y/n), elution at (65C/RT), dilution buffer (TE + RNAse/standard with NaCl and Triton X100 + RNAse)

2.25.1 other things to check

does a lower pH matter? should I resuspend the beads in the lower pH. would be nice to use that 96-well PCR strip magnet that also allows in place mixing

Chapter 3

Serial Analysis of Promoter Enrichment

THIS CHAPTER/PROJECT IS DEAD

See the first paragraph of this chapter for an intro to what I was trying to do. See the last paragraph of this chapter for info about why I stopped trying.

Wed Apr 19 11:49:18 EDT 2006

Chromatin Immunoprecipitation has become the standard *in vivo* method for determining whether a particular transcription factor binds a particular piece of DNA. ChIP-Chip, a techinique that combines ChIP with a tiling array, performs a similar task except the location doesn't have to be known in advance as all locations are tiled on a microarray. The sensitivity of ChIP-Chip in determining the binding site of a transcription factor is around +/- 500bp, and the accuracy of ChIP-Chip with noisy microarrays is unclear. Here, I propose to develop a new technique building on an earlier idea for quantifying gene expression termed: Serial Analysis of Gene Expression (SAGE). The new technique, Serial Analysis of Promoter Enrichment, is more similar to genetic footprinting approaches (Fig. 3).

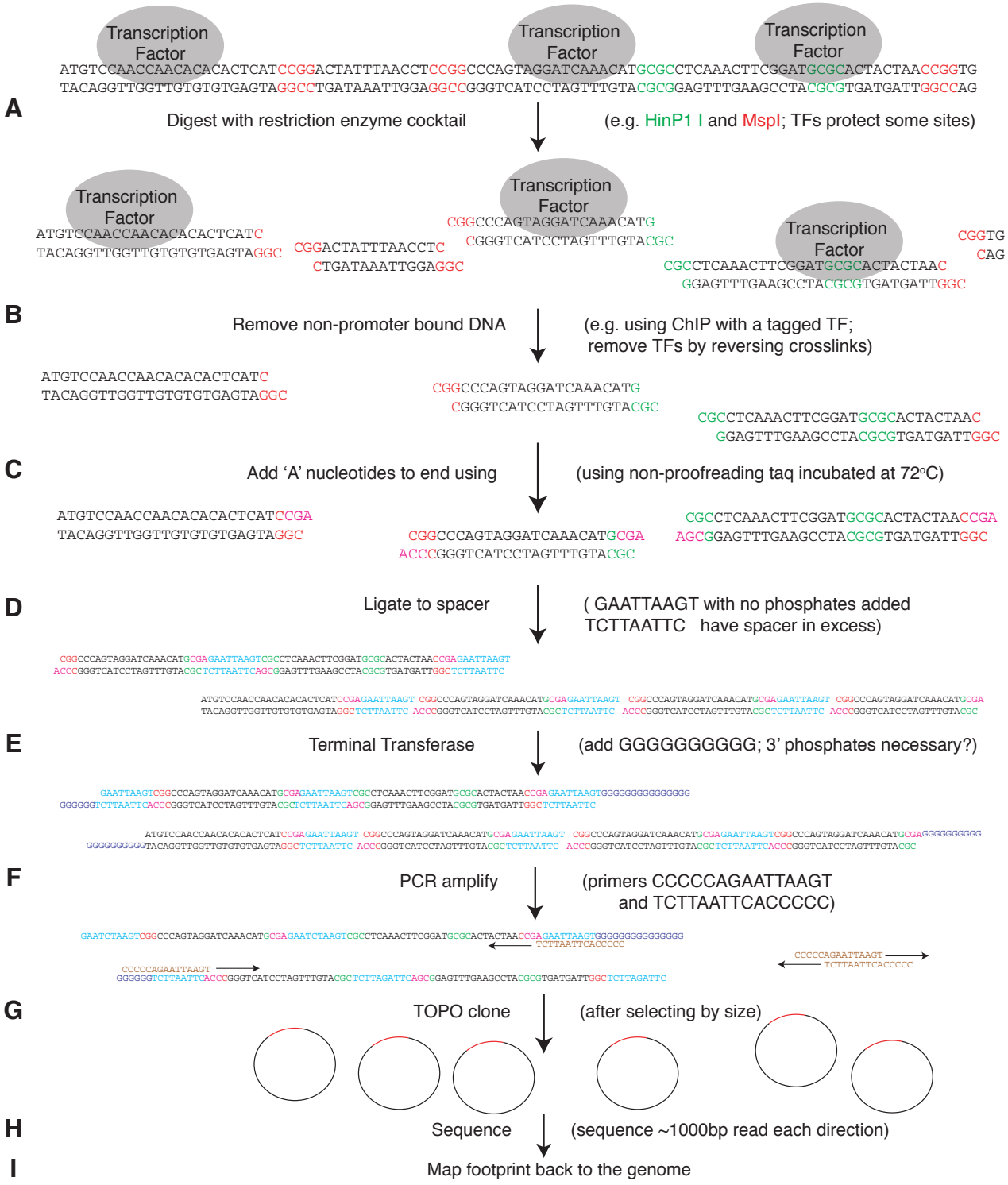
DNA is enriched *in vivo* or *in vitro* using a tagged transcription factor (as in ChIP). The bound DNA is digested with restriction enzymes or endonucleases. As with footprinting, the protein protects the binding site from digestion. Like in SAGE the fragments are ligated together with a separating linker, amplified, (size selected), and sequenced. Unlike SAGE the fragements are variable in length and the linker is connected in a different way.

3.1 Initial Steps

The linkers and primers have been chosen (Fig. 3). I'm going to try initially to digest a known plasmid (ignoring the ChIP part), glue the pieces back together, clone, sequence. This allows me to verify that the last steps are working when using a quantifiable amount of DNA and ratios (since ChIP DNA is not abundant enough to be quantified).

Here is some potentially useful info from IDT:

> **Annealing Protocol**
>
> It is sometimes necessary to make double-stranded DNA from single-stranded oligos. While the annealing procedure is very simple, attention to a few details can greatly reduce the presence of undesired single stranded material.
>
> Method:
>
> - Dissolve oligos in STE Buffer (10 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA). The presence of some salt is necessary for the oligos to hybridize. Dissolve each oligo at high concentration (1 - 10 OD260 units / 100 uL).
> - Mix two stands together in equal molar amounts. If you do not there will always be single stranded material left over.
> - Heat to 94oC and gradually cool. For many oligos this can be as simple as transferring to the bench-top at room temperature. For sequences with significant hairpin potential, a more gradual cooling/annealing step is beneficial; this is easily done by placing the oligos in a water bath or temp block and "unplugging the machine".
> - The resulting product will be in stable, double-stranded form and can be stored at 4oC or frozen.
>
> Things to consider: If the product will be used in a ligation reaction, the addition of 5' -phosphate may be needed. This can be done at the time of oligo synthesis (chemical phosphorylation) or at any time thereafter using PNK (enzymatic phosphorylation). If



the oligos are relatively long or to be used in cloning, starting with PAGE purified oligos is recommended.

from the FAQ

Can you tell me the minimum number of bases that can be annealed?
You can anneal an oligo of any size to its target, but longer sequences will lead to more stable duplexes. A minimum of 10 bases is needed for a PCR primer to find and remain annealed to its target long enough for extension to occur. This may be a good minimum threshold to consider if you plan to make dsDNA.

The initial linker has a melting temperature around 20C, slightly less than room temperature, but above the 16C optimum for ligation. It is also 8bp long (minus the 3' T hanging off each end), so it is shorter than the minimum needed for PCR primers and hopefully this will prevent any unligated linkers from making primer dimers during the amplification step.

3.1.1 First SAPE Protocol

Initially, I am removing the ChIP enrichment step. I know I can do that part. The tricky part is putting all the transcription factor binding sites into plasmids separated each with a linker. I'll start with a piece of known DNA (a plasmid), digest it, ligate it to the linkers, amplify it, and finally clone it. The **BOLD** capital letters for each step indicate the corresponding step in Figure 3.

1. digest 100 ng and 1000 ng of plasmid DNA with 0.5 and 5 units of EACH restriction enzyme in a 50 μl reaction for 45 min at 37C (MspI, MseI, HinP1 I) [5 units if 0.5 μl this is 5x the recommended amount]
2. ethanol precipitate
3. (**C**) add 20μl PCR master mix (do we really need to clean up the buffer after this?) incubate 20min at RT to add A's to the ends of each digestion product
4. (**D**) do *either* step 5, 6, 7, or 8
5. (**D**)
   (a) heat deactivate the restriction enzymes by incubation at 65C for 25 minutes
   (b) add ATP to a final conc of 1 mM
   (c) ligate to linker for different time lengths using 1 μl T4 ligase and linker at 16C
   (d) heat deactivate T4 at 65C for 10 min
6. (**D**)
   (a) clean up products with a Qiagen nucleotide removal or PCR clean up kit (NOTE: this removes the products below 40 and 100 bp respectively)
   (b) add ligase buffer
   (c) ligate to linker for different time lengths using 1 μl T4 ligase and linker at 16C
   (d) heat deactivate T4 at 65C for 10 min
7. (**D**)
   (a) clean up products by ethanol precipitation
   (b) add ligase buffer

(c) ligate to linker for different time lengths using 1 μl T4 ligase and linker at 16C

(d) heat deactivate T4 at 65C for 10 min

8. ethanol precipitate (we have to get rid of the ATP from the ligation)
9. (**E**) (for 10-20ng DNA) add 5 μl NEBuffer 5, 5 μl 2.5 mM CoCl2 soln, 0.5 μl Terminal Transferase enzyme, 1 μl 10mM dGTP incubate at 37C for 30 min
10. heat deactivate at 70C for 10 min
11. ethanol precipitate
12. (**F**) D amplify with the two PCR primers try a range of temps, lean on the high temp side to prevent the primer dimer problem; Min temp is 47C use easy-A kit
13. size select and gel purify 1-2 kb band
14. (**G**) TOPO clone
15. (**H**) sequence clones that have an insert

Finally getting started...

Fri May 12 12:25:40 EDT 2006

Unlike the above plan this is what I actually did.

1. miniprep 4 ml of lrpA and lrpB

Sample	DNA (ng/ul)	260/280	260/230
lrp A	110	1.93	1.28
lrp B	109.3	1.94	1.39

2. digest 2.5 ug plasmid DNA minipreped from lrpA and lrpB (in TOP10 cells); master mix of enzyme cocktail (0.25 μl MspI, 0.5 μl Mse, and HinPI; 30 μl H20, 10 μl BSA); digests were run in 50μl rxns for 1 hr at 37C followed by 20 min at 65C to heat deactivate the enzymes.
3. the ethanol precipitation was not done (step 2 above). rather 10 μl of water and 10μl NEB PCR master mix were added and the samples were incubated for 10 min at 72C to add the A's to the end. The ethanol precipitation was skipped because NEB Buffer 2 and the PCR buffer are fairly similar and I don't need an efficient amplification, I only need 2-3 bp filled in plus the A on the end.
4. ethanol precipitate to remove (hopefully) the Taq, dNTPs and the buffer; 7 μl NaAcetate (into the 70 μl), mix, add 160 μl 95% ethanol; -85C 20 min, spin 4C 10 min, vacuum out ethanol, add 750 μl COLD 70% ethanol, spin 4 min, (dry), resuspend in 40 μl TE
5. quantify DNA with the nanodrop

Sample	DNA (ng/ul)	260/280	260/230
lrp A	11.9	1.42	3.21
lrp B	24.8	1.65	2.45

16 μl of each were run on a 2% agarose gel (see Figure 3.1)

6. ligate; aliquoted 16 μl of lrpA and B (approx 190 ng and 396 ng of DNA respectively) and 2x 1.6 μl for both digests for a total of 6 ligation reactions; 3 of the tubes were ligated for 2 hr at 16C, the other 3 were ligated for 5 hrs. all six reactions were halted by 65C for 10 min.

7. ligations were cleaned up with a PCR purification kit. This kit removes DNA less than 100 bp which is ok because we don't want to clone or sequence those anyways, so we might as well not amplify them with the PCR. Ligated DNA was quantified

Sample	DNA (ng/ul)	260/280	260/230
lrp A 2hr L	10.2	3.53	0.70
lrp A 5hr L	9.9	4.95	0.67
lrp A 2hr H	11.8	3.94	0.65
lrp B 2hr L	11.0	3.52	0.67
lrp B 5hr L	10.1	4.5	0.82
lrp B 5hr H	12.5	3.47	1.04

*** Some strange things here *** No matter what my starting conc of DNA the amount of DNA in the ligation product is the same. All are very dirty too. ***

8. add GGGGGGGGGG tag with terminal transferase (TdT); 2 μl ligated DNA (post PCR purification), 5 μl Buffer 4, 5 μl CoCl2, 1 μl 100 mM dGTP, 0.5 μl TdT, 36.5 μl H2O.

9. amplified each sample via PCR; 12.5 μl NEB master mix, 1 μl 200 nM SAPE primer1 Forward and Reverse mix, 1 μl mix from TdT rxn above (not cleaned up. could this cause a problem? particularly the CoCl2?), 10.5 μl H2O; 10 μl of each PCR product was run on a 2% gel (see Figure 3.1). PCR annealing was at 48C. Maybe I should drop it, and raise the extension time (which as only 30sec at 72C, should make that much of a difference though).

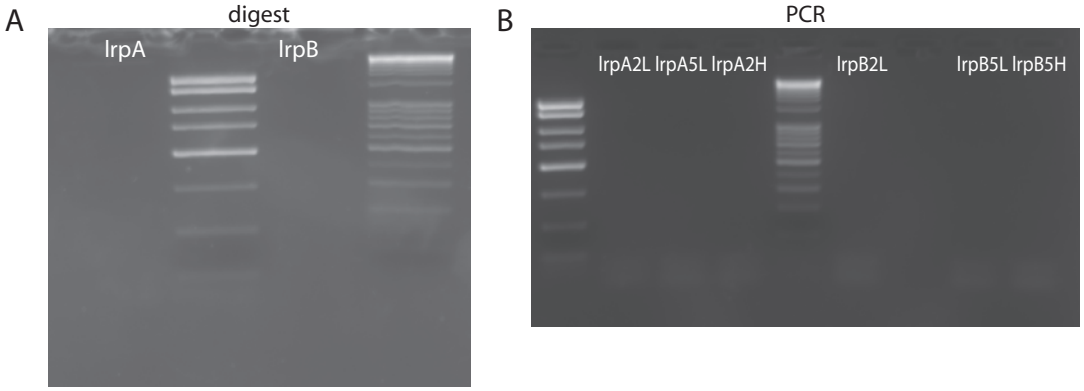


Figure 3.1: Diagnostic 2% agarose gels from first attempt at SAPE. The plasmid was digested for 1hr with three 4-mer cutters. The digested plasmid (A) seems so chopped up that I can't see any DNA on the gel (or perhaps it was lost in the ethanol precipitation?). The amplified PCR product didn't fair any better. There is a little something around 30bp but probably either primer dimer or just the primers.

Brief Conclusions: As is clear in Figure 3.1, the first round didn't work. Lower temp on PCR? Use less enzymes during digestion? (why the 10% DNA loss? why can't see it on gel?) Run ligation longer? ***Run ligated DNA on gel. This should help localize the problem (like a print statement

in debugging computer scripts). *** Try Tim's idea of attaching a linker with phosphorylated ends? Idea from Josh, use exonuclease to chop up single stranded stuff.

Trying to narrow down what went wrong in SAPE round one

Mon May 15 15:26:12 EDT 2006

Ran/running the ligated fragments for the three lrpA samples from step 7 above on a 2% agarose gel to see if there was anything at that stage. Ran 27μl of each (that's all I had). If the table for step 7 is accurate (and I don't believe it is) this should be about 275 ng of DNA. I initially pre-stained the gels with agarose. There appeared to be a very large band only visible in the tube that should presumably have the most DNA (based on starting material not based on the nanodrop reading in which is was the most abundant but only by a small margin). The band wasn't bright enough to concretely state that something was there. If it is correct, it appears the ligation step is definitely too long (that was only a 2hr ligation). I am not post staining the gel with sybr gold to see if it brightens the band a little. (I tried to remove a little of the EtBr by rotating the gel with H2O for 15 min with a water change after the first 10 min). The sybr gold is being rotated with the gel for 25 min.

There does seem to be a very faint LARGE band (see Figure 3.2) at around 10000bp. Unfortunately the agarose gel percent chosen is horrible for separating at this range, so its hard to really say what size it is. The band is slightly more visible with Sybr Gold than EtBr (I think our Sybr Gold is getting old). You really have to zoom in a lot to see the band (700% or so).

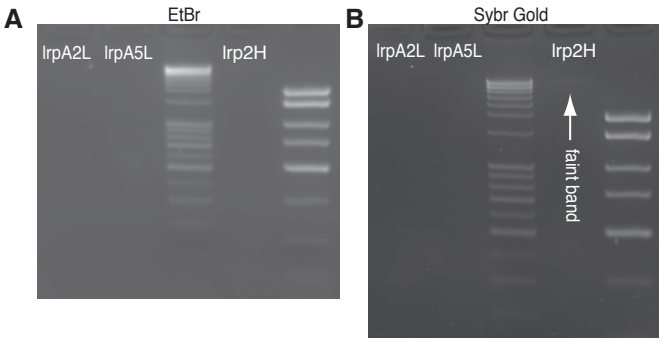


Figure 3.2: Diagnostic 2% agarose gels from first attempt at SAPE. 27μl of the qiagen PCR cleaned-up ligation reaction was run on the gel. The gel was originally pre-stained with EtBr (A). Subsequently it was washed 10 min and 5 min with water and finally dyed with 1x Sybr Gold for 25 min and washed for 10min in H2O (B). The faint band shows what seems to be an appx 10kb band of the ligated product.

Brief Conclusions: Figure 3.2 seems to indicate the ligation was too long. I'm not sure what to conclude about why this PCR didn't work based on this. It could either be 1) the ligation products were too long to allow such a long amplification (particularly given that my amplification time was only 30 seconds) 2) the end labeling with poly-G didn't work 3) both

Trying to further narrow down what went wrong in SAPE round one

Tue May 16 12:25:13 EDT 2006

I'm running out the lrpB ligations on a gel as I did for lrpA yesterday. I am *only* going to stain with Sybr Gold, so hopefully I'll get a better signal. Used 7.5 μl of ladder instead of the standard 10 μl so it doesn't overwhelm the weaker signals. I also included 110 ng and 109 ng of lrpA and lrpB miniprep product respectively for size comparison and a sanity check just to make sure the minipreps all of these results are dependent on really worked.

I really think the Sybr gold is going bad. I used the solution from yesterday and the gel was completely blank. I tried switching to EtBr and could barely see the bands. Now I increased the concentration of Sybr Gold (after so much watching I don't know if I have any DNA left in the gel anymore).

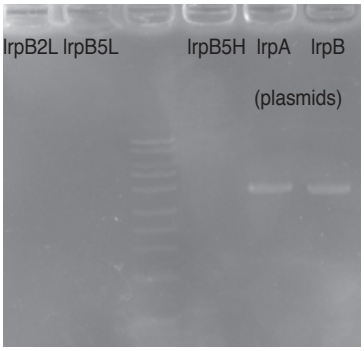


Figure 3.3: Diagnostic 2% agarose gels from first attempt at SAPE. 27μl of the qiagen PCR cleaned-up ligation reaction was run on the gel. The gel was originally pre-stained with Sybr Gold. Subsequently it was stained with EtBr. Subsequently it was stained with Sybr Gold again (but much more concentrated). I think the old Sybr Gold killed this gel. The two lanes of plasmid DNA are the correct size

Brief Conclusions: This didn't work terribly well (see Figure 3.3). Mainly a problem with the gel staining. Glad to at least see the two plasmids were there.

SAPE1 try 2 (probably without going to the PCR step)

Tue May 16 12:28:03 EDT 2006

I'm trying the initial steps again, this time with shorter ligation times. I'm also using fewer restriction enzymes (only MspI) so hopefully I'll be able to see the digest DNA better (and I'll use Sybr Gold to help that problem too). Below is the sizes of the pieces I should theoretically get (based on an *in silico* digestion with NebCutter2).

Fragment	Cutter Ends	Coordinates	Length (bp)
1	MspI-MspI	4048-208	551
2	MspI-MspI	2531-3057	527
3	MspI-MspI	334-834	501
4	MspI-MspI	869-1310	442
5	MspI-MspI	1764-2167	404
6	MspI-MspI	3307-3698	392
7	MspI-MspI	1311-1552	242
8	MspI-MspI	3699-3939	241
9	MspI-MspI	2168-2357	190
10	MspI-MspI	3092-3239	148
11	MspI-MspI	2384-2530	147
12	MspI-MspI	209-333	125
13	MspI-MspI	1553-1662	110
14	MspI-MspI	3940-4047	108
15	MspI-MspI	1663-1729	67
16	MspI-MspI	3240-3306	67
17	MspI-MspI	835-868	34
18	MspI-MspI	1730-1763	34
19	MspI-MspI	3058-3091	34
20	MspI-MspI	2358-2383	26

For the digests, all of the remaining DNA was used (around 2.5 μg). The digestions were done with MspI *only* for 1 hr followed by addition of 10 μl H_2O and 10 μl PCR Master mix [NEB]. After ethanol precipitation, the samples were resuspended in 20 μl TE. lrpA was immediately run on a 2% gel (Figure 3.4A). The lrpB digestion was ligated for 30 min and then the entire ligation reaction was run on a gel (Figure 3.4B).

DNA was quantified after the digestion was cleaned up via ethanol precipitation:

Sample	DNA (ng/ul)	260/280	260/230
lrp A digestion	71.2	1.62	2.68
lrp B digestion	31.1	1.81	3.25

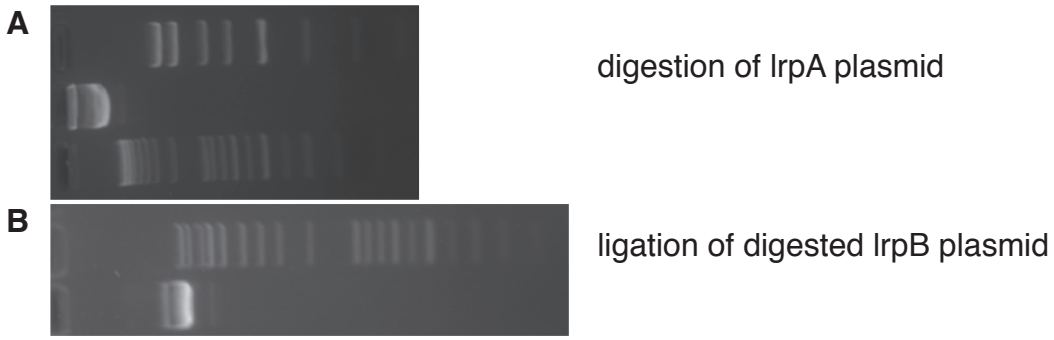


Figure 3.4: (A) is 1% (B) is 2%

Brief Conclusions: The DNA yield post-digestion cleanup were much better though this may be because the previous elution volume was 40 μl which was too dilute for reliable measurement.

I think the problem with SAPE1 is clear now. Tomorrow, I'll run the digestion again with out the addition of the Taq at 72°C for 10 min. If the digestion looks correct, then the puzzle is solved. It appears that all the little pieces from the digestion bind to each other when in the solution with the Taq at 72°C and prime the PCR reaction resulting in all the pieces being turned into a series of long giant pieces (see Figure 3.4). It's kinda puzzling to me why NO smaller pieces appear though.

SAPE1: solving the puzzle

Wed May 17 16:12:35 EDT 2006

Three replicates of lrpB were miniprepped with the following yields:

Sample	DNA (ng/ul)	260/280	260/230
lrp B 1	82.5	1.99	2.13
lrp B 2	70.4	2.00	2.20
lrp B 3	80.6	1.98	2.03

I took a 24 μl sample from lrpB1 and lrpB2 and digested them with MspI (0.25 μl) and HinP1 I (0.5 μl) respectively. I ran the entire digestion on a gel (unfortunately I forgot to run a little of each plasmid out as the same time Figure 3.5).

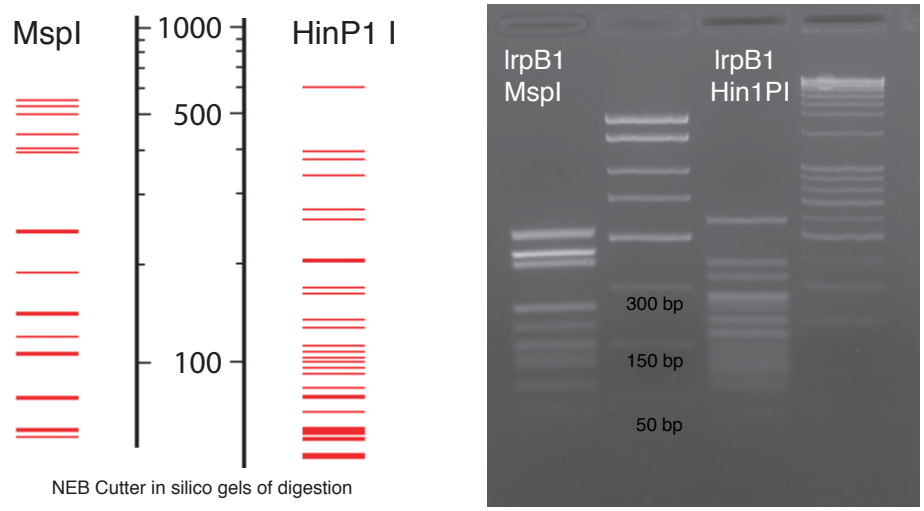


Figure 3.5: *in silico* and *in vitro* digests of the lrpB plasmid with 4-mer cutters MspI and Hin1PI without the addition of PCR master mix which messes everything up.

Brief Conclusions: So the problem is now found (but not solved at least not experimentally yet). Figure 3.5 shows that the digestions are working correctly this leaves only the Taq based addition of A's (step 3 of SAPE **First SAPE Protocol**) as the culprit for the huge pieces of DNA found in Figures 3.2 and 3.4. TA based SAPE will not work without something clever. I plan to abandon TA based SAPE in SAPE2.

3.1.2 SAPE1: the return of SAPE1

Fri May 26 11:31:32 EDT 2006

Jay Shendure did a technique almost identical to what I want to try in George Church's lab in *Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome* a recentish Science paper. His situation was sheared DNA with messy ends. He used the End-It DNA Repair Kit [Epicenter] to make the ends blunt. Then he added taq with ONLY dATP to put the A on the end. I'm going to try this and see if it fixes the huge product problem (see Figure 3.4)

1. digest 24 μl plasmid DNA minipreped from lrpB1 and lrpB2; master mix of enzyme cocktail (0.25 μl MspI, 5μl Buffer 2, 21 μl H_2O); digests were run in 50μl rxns for 1 hr at 37C followed by 20 min at 65C to heat deactivate the enzymes.

2. ethanol precipitate; 5 μl NaAcetate (into the 50 μl), mix, add 110 μl 95% ethanol; -85C 20 min, spin 4C 10 min, vacuum out ethanol, add 750 μl COLD 70% ethanol, spin 3 min, (dry) [1]

3. resuspend in 34 μl TE

4. add 5 μl End-Repair Buffer, 5 μl dNTP Mix, 5 μl 10mM ATP, 1 μl End-Repair Enzyme Mix

5. incubate RT for 45 min, deactivate at 70 C for 15 min

6. run all of lrpB3 on a gel to make sure it is ok, post End-Repair (put lrpB3 in freezer, so all could be run on the same gel)

7. ethanol precipitate; 5 μl NaAcetate (into the 50 μl), mix, add 110 μl 95% ethanol; -85C 20 min, spin 4C 10 min, vacuum out ethanol, add 750 μl COLD 70% ethanol, spin 3 min, (dry) [2]

8. resuspend in 25 μl TE

9. A-tail DNA: 5 μl standard taq (mg-free) rxn buffer [NEB], 6 μl 25 mM MgCl2 (3 mM final), 0.25 μl dATP (0.5 mM final), 0.5 μl Taq (2.5 units), 25 μl sample

10. run all of lrpB2 on a gel to make sure it is ok, post A-tail (put in freezer, so all could be run on the same gel)

11. ethanol precipitate; 5 μl NaAcetate (into the 50 μl), mix, add 110 μl 95% ethanol; -85C 20 min, spin 4C 10 min, vacuum out ethanol, add 750 μl COLD 70% ethanol, spin 3 min, (dry)

12. resuspend in 16 μl TE

13. ligate; 16 μl of lrpB3 for 2 hr at 16C. halted by incubation 65C for 10 min. (1 μl SAPE1 linker, 16 μl sample, 2 μl buffer, 1 μl T4 DNA ligase [NEB])

14. run lrpB3 on a gel (again put in freezer)

Sun May 28 18:35:09 EDT 2006

All three samples (lrpB1,2,3) are being run on a 1.5% agarose. In addition, I made two midipreps using the C.6.4 on page 425[3], so I should have lots of plasmid and not have to waste a bunch of money and time on the Qiagen preps. One of the samples was also run on the gel. Yields were:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**
lrp B 1 Midi	1613.4	2.14	2.36
lrp B 2 Midi	1779	2.14	2.37

[1] I spilled a little of lrpB3 here :(Only a few microliters though.

[2] had strong white precipitate. would assume to be DNA but wasn't in previous etOH prep. I might need a phenol/chloroform step to prevent this :(

[3] unlike that protocol I used EC broth because I was out of LB

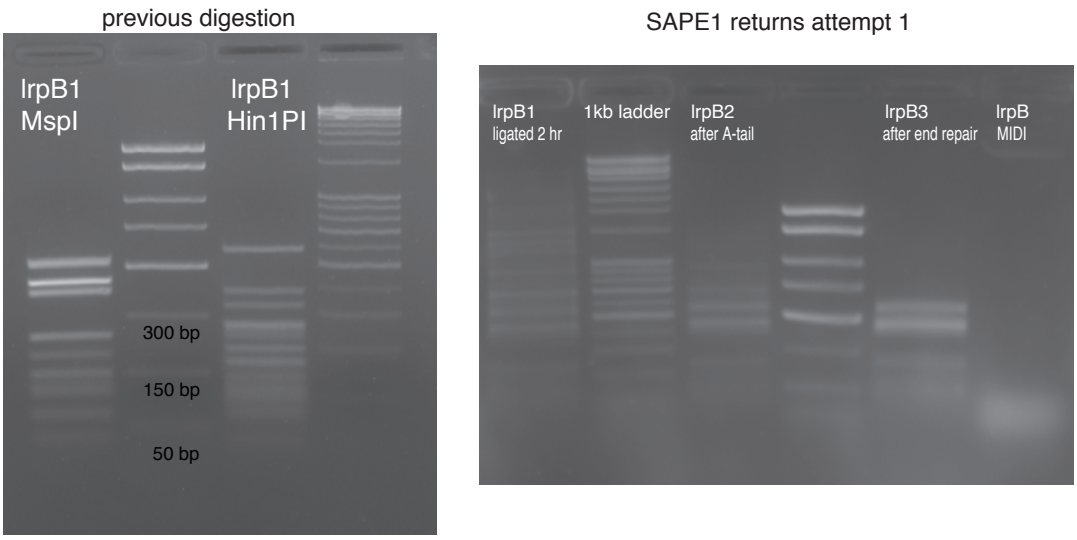


Figure 3.6:

Brief Conclusions: I wish I started with a little more DNA. It would be nice if the DNA were concentrated enough that I could track a single sample through all the steps (rather than assuming that lrpB1 is like B2 is like B3). I shouldn't have switched to a 1.5% gel (see Figure 3.6 right pane). The lower wt bands are hard to see. Even the ladder is hard to see as far as that goes. My gels seem to be getting crappier recently. However, I think that ligation is working. Unfortunately it is hard to see the lrpB3 lane (again I need to start with more DNA). The lrpB2 post A-tail where I used to have problems looks a lot better. It does start to get a little smeary towards the top, but not the giant piece like I was getting previously (e.g. Figure 3.4). And the best thing is the DNA starts to get larger in size after the ligation step which is just right. I think adding a phosphatase step before the A-tailing would be good. I don't think I can conclude if lrpB3 sucks because the end-repair messed it up or because I spilled the tube. Ideally I'd run one sample straight through starting with about 5x my normal starting amount of DNA (e.g. 12.5 μg) and pull out a sample to run after: 1) digestion, 2) end-repair, 3) A-tail, 4) ligation. And maybe 200 ng of uncut plasmid just to be safe. The MIDI prep using the new protocol does not seem to have worked (Figure 3.6). I need this to work if I'm going to have enough DNA.

Midipreps

Mon May 29 2006

I need more DNA if I'm going to trace one sample all the way through this process (I think I need around 10 μg of DNA). I tried to do a midiprep using the protocol in *Molecular Cloning*. My version of the protocol (little to no modifications just clarified things based on my previous attempt above) can be found on page 425. I ran 4 preps using 15 ml of culture for each. The results of each of these four and the one from the previous day that I didn't run on a gel are in Figure 3.7. One microliter was run in each lane (combined with 9 μl H_2O and 1.6 μl 6x loading dye). Yields were very high:

Sample	DNA (ng/ul)	260/280	260/230
lrp B 1 Midi	2073	2.12	2.34
lrp B 2 Midi	1679.6	2.13	2.36
lrp B 3 Midi	1720.9	2.13	2.35
lrp B 4 Midi	2215.6	2.12	2.34

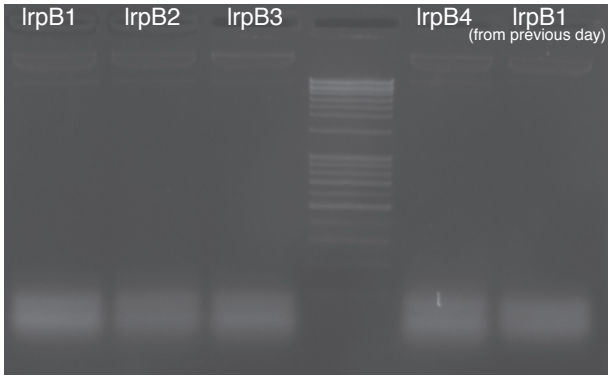


Figure 3.7:

Brief Conclusions: This sucks. I seem to have only gotten a large amount of tiny fragments in both of my attempts at midipreps (see Figures 3.6 last column and 3.7). I'm giving up on this protocol for now. My only guess as to what went wrong is that the phenol:chloroform is old, but I don't know if that stuff goes old. The bottle is almost empty so I ordered new and will just toss the remaining 3 ml or so. For now I ordered the Qiagen HiSpeed Midi kit. Hopefully that will give me more success than the old school Sambrook method.

3.1.3 SAPE1 returns: following the same sample all the way (including PCR)

Wed May 31 14:45:54 EDT 2006

I won't get started until Mon cause sis and mom are in town, but want to build my gameplan for the week. Gameplan broadly:

1. inoculate 3 x 50 ml cultures for monday (Sunday)
2. make a ton of DNA with the midiprep kit for 3 samples (Monday)
3. test the 58 operon promoters alone and perhaps in couples with NEB master mix (Monday)
4. run SAPE1 to completion test samples at each step include G tail and PCR (Tuesday)
5. try SAPE2 (Wednesday)

The plan for SAPE1

Wed May 31 15:32:25 EDT 2006

Modify below based on conc of DNA at each step.

1. digest 24 μl plasmid DNA midipreped from lrpB1 and lrpB2; master mix of enzyme cocktail (μl MspI, 5μl Buffer 2, 21 μl H_2O); digests were run in 50μl rxns for 1 hr at 37C followed by 20 min at 65C to heat deactivate the enzymes.
2. run 1/5 of sample on gel
3. ethanol precipitate; 5 μl NaAcetate (into the 50 μl), mix, add 110 μl 95% ethanol; -85C 20 min, spin 4C 10 min, vacuum out ethanol, add 750 μl COLD 70% ethanol, spin 3 min, (dry)
4. resuspend in 34 μl TE
5. add 5 μl End-Repair Buffer, 5 μl dNTP Mix, 5 μl 10mM ATP, 1 μl End-Repair Enzyme Mix
6. incubate RT for 45 min, deactivate at 70 C for 15 min
7. is Antarctic phosphatase step possible. try adding antarctic phosphatase buffer to 1x. calculate units based on conc of DNA. incubate 37 C for 30 min, heat deactivate 65C for 5 minutes
8. run 1/4 of sample on gel
9. ethanol precipitate; 5 μl NaAcetate (into the 50 μl), mix, add 110 μl 95% ethanol; -85C 20 min, spin 4C 10 min, vacuum out ethanol, add 750 μl COLD 70% ethanol, spin 3 min, (dry) [4]
10. resuspend in 25 μl TE
11. A-tail DNA: 5 μl standard taq (mg-free) rxn buffer [NEB], 6 μl 25 mM MgCl2 (3 mM final), 0.25 μl dATP (0.5 mM final), 0.5 μl Taq (2.5 units), 25 μl sample
12. run 1/3 of sample on gel
13. ethanol precipitate; 5 μl NaAcetate (into the 50 μl), mix, add 110 μl 95% ethanol; -85C 20 min, spin 4C 10 min, vacuum out ethanol, add 750 μl COLD 70% ethanol, spin 3 min, (dry)
14. resuspend in 16 μl TE
15. ligate try a different ligation time for each of the samples; 16 μl of lrpB3 for 2 hr at 16C. halted by incubation 65C for 10 min. (1 μl SAPE1 linker, 16 μl sample, 2 μl buffer, 1 μl T4 DNA ligase [NEB])
16. run 1/2 of sample on gel
17. end label with dGTP
18. PCR amplify with some of the G-labeled DNA
19. run PCR product to see if it worked

[4] had strong white precipitate. would assume to be DNA but wasn't in previous etOH prep. I might need a phenol/chloroform step to prevent this :(

3.1.4 More DNA: midipreps, midipreps, midipreps

Wed Jun 7 17:42:23 EDT 2006

I've been having a heck of a time getting midipreps to work. I tried an old-fashioned approach and a the new HiSpeed Qiagen midiprep. The Qiagen kit gave me clean DNA but not very much of it (see 3.8 and table below). I'm not sure why the yield was so low. I haven't yet tried Qiagen's debugging process. Meanwhile I figure out the problem with the earlier midipreps and the huge amount of crap right around 50 bp (see Figures 3.7 and 3.9A). The problem was that adding RNA too the 1st lysis buffer was insufficient to get rid of all the RNA it seem to chop it into this small fragment size. I read on the web a good place for the RNA digestion is right after you precipitate the cellular wall and genomic DNA (right before the phenol chloroform step). In Figure 3.9, I digested the midiprep with a large amount of RNAse cocktail and RNAse A (sample on the right only) and the RNA the samples (before RNAse Figure 3.9A). The RNA was also drastically messing up my measurements of DNA concentration.

Yields from Qiagen midipreps (500 μl total volume; run on Mon Jun 5, 2006):

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	total yield (ug)
1	32.6	1.76	2.42	16.3
2	77.1	1.62	0.85	38.6

Yields from old-school midipreps (100 μl total volume; run on Tue Jun 6, 2006):

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	total yield (ug)
1	3838.2	2.02	2.19	383.8
2	3727.8	2.04	2.21	372.8

Yields from old-school midipreps post RNA digestion (100 μl total volume; quantified Thur Jun 8, 2006):

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	total yield (ug)
1	72.3	1.90	2.22	7.2
2	63.3	1.88	2.17	6.3

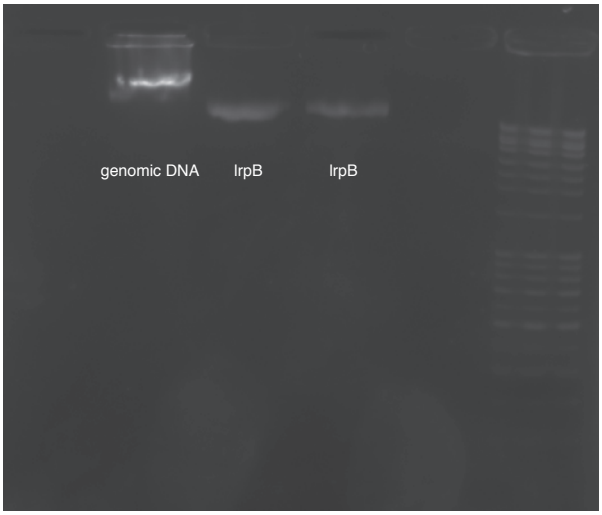


Figure 3.8: Qiagen midipreps didn't produce nearly as high a yield as they claim I can get in the manual. Genomic DNA in the first lane is for use in the operon study.

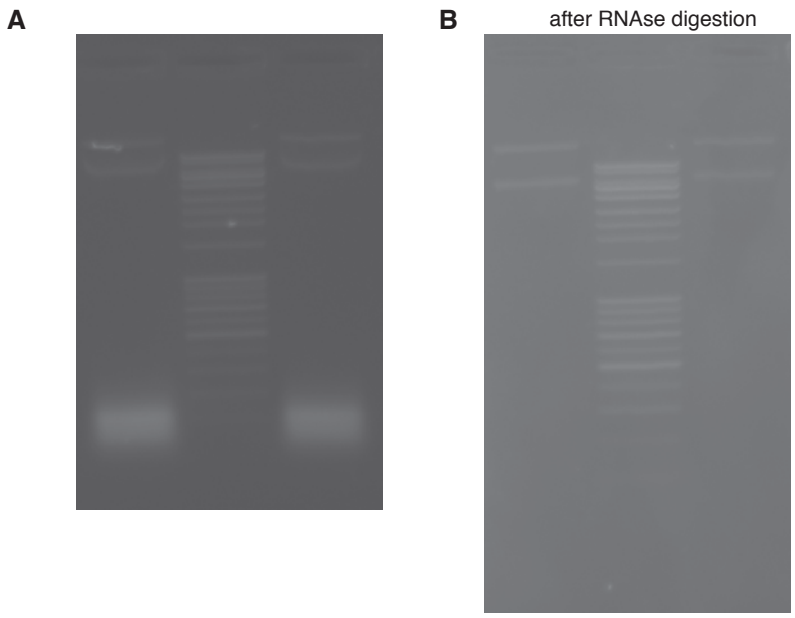


Figure 3.9: 0.75 μl of midiprep DNA (total volume) from the old-school method on a 1% agarose gel. Initially RNA was high and yield was high. Then I digested with RNAse cocktail and the RNA went away (as did the yield which I guess was mostly RNA).

Mon June 12, 2006 Ran six more Sambrook mini-preps with only slightly more luck than before. Here are the yields before RNAse digestion.

Sample	**DNA (ng/ul)**	**260/280**	**260/230**
lrp B 1 Midi (10 ml starting culture)	670	1.93	2.18
lrp B 2 Midi (10 ml starting culture)	1160.8	2.10	2.34
lrp B 3 Midi (15 ml starting culture)	950.9	2.08	2.34
lrp B 4 Midi (15 ml starting culture)	1246.6	2.05	2.34
lrp B 5 Midi (25 ml starting culture)	1935.0	2.10	2.29
lrp B 6 Midi (25 ml starting culture)	1519.8	2.09	2.32

and after RNAse digestion.

Sample	**DNA (ng/ul)**	**260/280**	**260/230**
lrp B 1 Midi (10 ml starting culture)	186.7	1.81	1.57
lrp B 2 Midi (10 ml starting culture)	284.5	1.82	1.73
lrp B 3 Midi (15 ml starting culture)	246.1	1.84	1.66
lrp B 4 Midi (15 ml starting culture)	237.2	1.84	1.34
lrp B 5 Midi (25 ml starting culture)	246.2	1.78	1.33
lrp B 6 Midi (25 ml starting culture)	152.9	1.75	1.27

3.1.5 Midipreps that work and are RNA free!

Sun Jun 18, 2006

I was just about to give up on the midi idea altogether. I grew cells to make 6 Qiagen minipreps (with 5 ml of culture apiece to get maximal yield). I also included a new protocol for midipreps (see

section C.6.3 page 423). I planned to ethanol precipitate the 6 preps into one more concentrated prep if the midi didn't work.

Yields were as follows:

Sample	DNA (ng/ul)	260/280	260/230
lrp B 1 Mini (5 ml starting culture)	163.9	1.92	2.10
lrp B 2 Mini (5 ml starting culture)	155.5	1.93	2.09
lrp B 3 Mini (5 ml starting culture)	143.2	1.90	1.86
lrp B 4 Mini (5 ml starting culture)	150.5	1.92	1.97
lrp B 5 Mini (5 ml starting culture)	146.7	1.92	2.03
lrp B 6 Mini (5 ml starting culture)	159.1	1.92	2.02
lrp B 1 Midi (50 ml starting culture)	395.5	1.94	2.30

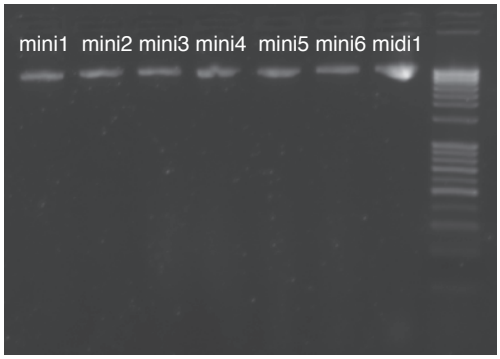


Figure 3.10: 0.75 μl of miniprep (6 samples) and midiprep (1 sample) DNA (total volume) from on a 1% agarose gel. RNA is not visible. DNA is not nicked for the Midiprep as it has been with the Sambrook Midiprep.

Brief Conclusions: After much frustration getting a simple technique to work, I finally found a protocol that works. And it seems to work well (see Figure 3.10). Unlike with the Sambrook protocol midipreps, this new method doesn't result in nicked DNA (not sure why). Unfortunately, I only had two midiprep samples and one of them was spilled before I completed the protocol. But for now, I'll assume this annoying search for a plasmid prep with high yield and RNA free DNA is over.

3.1.6 SAPE2 a modified protocol that hopefully works (original SAPE did not)

Quick thought on SAPE1...Maybe it didn't work because I didn't clean up the reaction before adding the Taq? I'd like to check that out using the method from Shendure in the Church lab to add ends (actually I have see section 3.1.2).

3.1.7 stab at an initial SAPE2 detailed protocol

Wed May 31 15:32:57 EDT 2006

Stab at a first protocol.

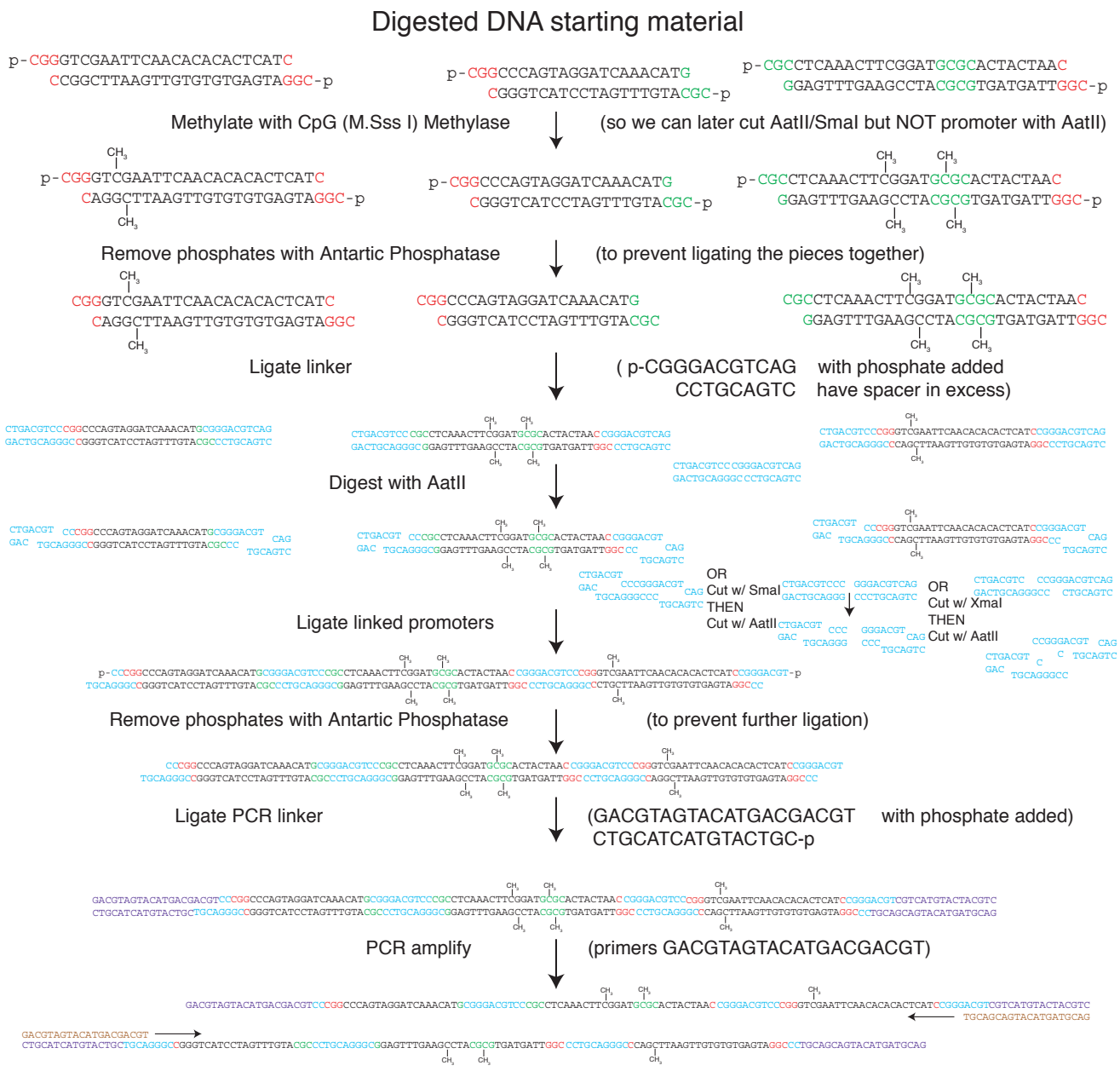


Figure 3.11: If one primer PCR is a problem, we can try the end labeling trick from SAPE1

1. test Sss I by running on a smaller sample and verifying that it blocks MspI and AatII digestion (next step)

2. digest 24 μl plasmid DNA midipreped from lrpB1 and lrpB2; master mix of enzyme cocktail (μl MspI, 5μl Buffer 2, 21 μl H_2O); digests were run in 50μl rxns for 1 hr at 37C followed by 20 min at 65C to heat deactivate the enzymes.

3. SssI protection of CG sites; heat inactivation 20 min at 65 C

4. add Antarctic phosphatase buffer and enzyme to remove phosphates from the end 15 min 37

C (is 15 min enough?), deactivate 65 C for 5 min

5. run 1/5 of sample on gel
6. ethanol precipitate; 5 μl NaAcetate (into the 50 μl), mix, add 110 μl 95% ethanol; -85C 20 min, spin 4C 10 min, vacuum ou t ethanol, add 750 μl COLD 70% ethanol, spin 3 min, (dry)
7. resuspend in μl TE
8. ligate phosphorylated linker in NEB buffer for with 1 mM ATP added. heat inactivate ligase
9. digest with AatII (do I need to ethanol precipitate?). or digest with SmaI then with AatII
10. ligate linked promoters (maybe add more dATP in)
11. (stop for now run on a gel) don't have primers for the rest of this. make sure this works first
12. remove phosphates with Antarctic Phosphatase
13. finally get to use a freaking kit. before PCR amplification, clean up DNA with Qiagen PCR cleanup kit

3.2 The end of SAPE

Wed Aug 23 15:19:15 EDT 2006

I think I'm going to abandon SAPE. The paper: *Multiplex sequencing of paired-end ditags (MS-PET) a strategy for the ultra-high-throughput analysis of transcriptomes and genomes.* pretty much does what I wanted to do, but maybe even better (though there protocol is no easier than mine and perhaps harder). I think it would be more effective to make PETs like I'm doing with my experimental transcript determination study. Then I could just sequence them to determine binding sites.

Chapter 4

chromatin immunoprecipitation based determination of transcription factor binding sites: SAPE returns

THIS CHAPTER/PROJECT IS IN PROGRESS

In Chapter 3, I sketched out some ideas and ran some preliminary experiments to try and achieve high-throughput transcription factor target discovery by adapting ideas from SAGE (serial analysis of gene expression) and DNA footprinting to ChIP (chromatin immunoprecipitation). However, my ChIP protocol at the time was quite tedious and I found a similar project had already been completed by another, so it was no longer necessary to publish yet-another-ChIP-pilot study.

With my recent factorial and response surface optimizations of the ChIP protocol (Chapter 2), the determination of all of the binding sites for all of the transcription factors in *E. coli* looks like an attainable goal. In addition, highly parallel sequencing has really taken off in the last year, so finding a place to outsource the sequencing should be easier.

The general strategy for determining the transcription factor binding sites will be to run a standard ChIP protocol (using one of my optimized protocols) but with an additional digestion step to shorten the length of the transcription factor-bound fragments. Finally, I'll purify the DNA fragments, barcode them, sequence them, and map them back to the genome (Figure 4.1).

4.1 Will restriction enzymes digest crosslinked DNA

Tue Nov 27 19:26:01 EST 2007

A useful component to generating an *in vivo* footprint is the ability to digest or cut DNA. The factorial ChIP protocol optimizations I've done, yielded one protocol that uses only 0.1% formaldehyde (see Chapter 2 for details and http://blog-di-j.blogspot.com/2007/12/optimized-chip-protocols.html for the optimized protocols. I've been told by a number of folks that it's difficult or not possible to generate a footprint with ChIP because crosslinked DNA is not readily digestable by restriction enzymes.

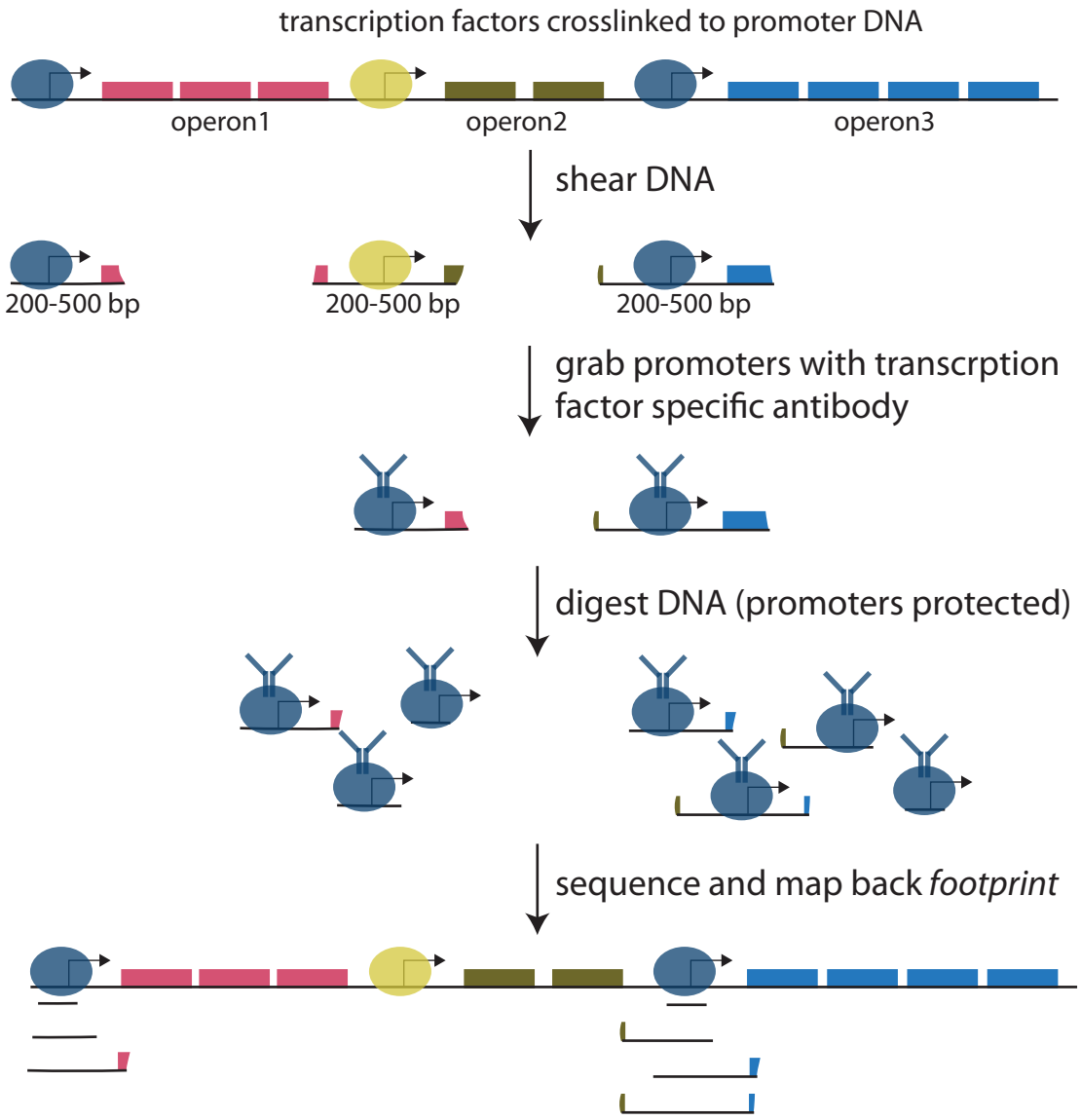


Figure 4.1: A schema for highly-parallel discovery of transcription factor binding sites.

For this experiment, I'm going to cut pUC19 (my favorite piece of DNA) with a bunch of restriction enzymes (mostly 4-mer cutters) to see how well they cut crosslinked DNA.

I crosslinked the DNA using 0%, 0.1%, or 1% formaldehyde for 10 minutes at RT followed by glycine quenching just like I do with my standard ChIP protocol. I used 7 μg of pUC19 in a total volume of 200 μl TE for each formaldehyde concentration. I added 0.74 μl formaldehyde and 7.4 μl formaldehyde for the 0.1% and 1% solutions respectively. All 3 samples were quenched with 10 μl of 2.5M glycine.

I EtOH precipitated all 3 samples with 1 μl of glycoblue and washed them all 2x with 750 μl of 70% EtOH. I resuspended each sample into 40 μl TE.

I split all 3 samples into 7 for the digestions. Each sample used: 0.5 μl enzyme, 2 μl BSA, 2 μl

NEBuffer, 5 μl pUC19 (from the 40 μl total), and 10.5 μl H_2O . I used either buffer 2 or buffer 4 (whichever was better for the particular enzyme). The seven tested enzymes were: HinPI, MseI, MspI, SphI, BamHI, NlaIII, BfuCI. All were incubated at 37C for 15 minutes and run directly on a 1.5% agarose gel (Figure 4.2).

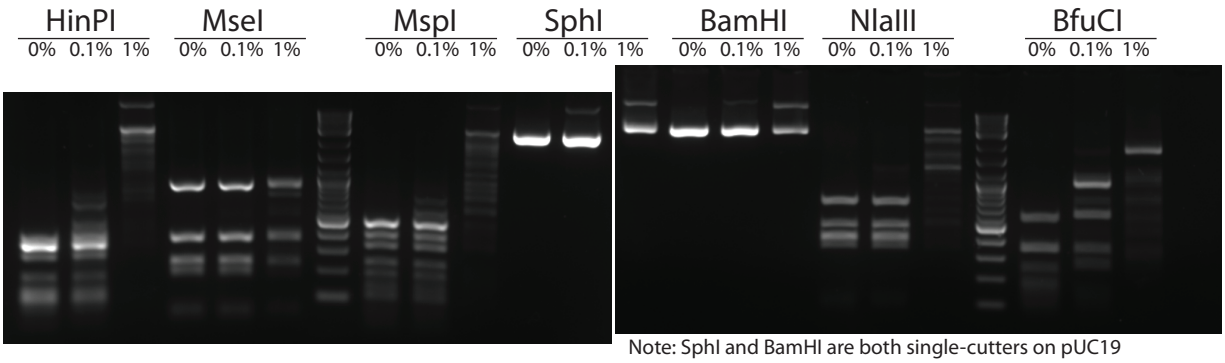


Figure 4.2: gel of cut pUC19 with different enzymes and different concentrations of formaldehyde used for crosslinking

Brief Conclusions: 1% clearly sucks and is hardly cuttable as I've heard from other folks. However, 0.1% formaldehyde cuts almost as well as the 0% (no crosslinking) Figure 4.2, so given that the 1.5 day ChIP protocol works with this amount, we definitely have some potential to trim our ChIP DNA (probably in between the TE washes) to obtain something footprint-like. It will hopefully also make it easier to clone or circularize (for RCA) the ChIP DNA. If circularization is easier, could be very useful for ChIP-Chip without needing LM-PCR (hopefully cleaner and easier). Though, I'm more interested in using this for ChIP-Seq.

random thoughts:

might want to digest with a large cocktail (or multiple different cocktails) of enzymes

digesting for different times might allow something more of a distribution of sites that are centered around the binding site instead of just multiple clones of the binding site (which could be too small to map back?)

4.2 Cloning ChIP DNA

Given the result above that shows the 0.1% formaldehyde used in the 1.5 day protocol is cuttable with restriction enzymes, I've decided to try and clone the DNA into pUC19 so I can sequence it.

4.2.1 preparing the enriched/cut Chromatin

Dec 5, 2007

I ran the 1.5 day ChIP protocol Version 1.0 using samples 1 (lrp) and 4 (pdhR) from section 2.19 on page 169. As in that section, I used 16 μl sheared chromatin with 64 μl of dilution buffer. I used 2.7 μl antibody (3.3 μg) and 100 μl beads. The only alteration to the protocol was a cutter-wash step in 100 μl of cutter mix after the two TE washes.

Each 100 μl of cutter mix contained: 0.5 μl BfuCI, 0.5 μl NlaIII, 10 μl NEBuffer4, 10 μl BSA, 79 μl H_2O . The beads + cutter mix was incubated at 65C for 15 minutes prior to eluting via the standard dynal elution procedure.

Dec 6, 2007

The two samples each with +/- antibody (for tubes total) were cleaned up with a Qiagen PCR purification kit. Unlike the typical ChIP procedure where I elute into 100 μl for downstream PCR, this time I wanted the DNA more concentrated for cloning. I eluted into the minimum volume of 30 μl .

DNA yield

Dec 6, 2007

Since the DNA was a little more concentrated than usual, I decided to see if I could concentrate it for the first time. I used 10 μl of the 30 μl sample with the Qubit hsDNA dye [Invitrogen]. The yields were:

sample	**yield**
lrp 1 (with antibody)	1.356 ng
lrp 1 (without antibody)	too low to measure
pdhR 4 (with antibody)	0.612 ng
pdhR 4 (without antibody)	too low to measure

Brief Conclusions: It's cool to see my ChIP DNA yield for the first time! I'm also quite happy to see that when I don't use antibody, I don't pull down a measurable amount of DNA (though the PCR reactions strongly suggest I do have some DNA in my no antibody control).

4.2.2 cloning the cut chromatin

Dec 6, 2007

I cut 1 μg of pUC19 cloning vector with each of the follow combinations of restriction enzymes (three samples total; one for each combination): BamHI (buffer 3 + BSA); SphI (buffer 2); BamHI + SphI (buffer 3 + BSA). These three vector cuts should allow me to capture any of the possible cut combinations on my ChIP DNA. I used 10 μl of each ChIP sample (I only used the two samples that contained the antibody not the negative controls), 2 μl ligase buffer, 5 μl H_2O , 2 μl vector mix (all three cuts mixed together; 3 ng total), and 1 μl T4 DNA ligase.

I ran standard heat shock tranformation and plated 75 μl of the 300 μl of cells on an ampicillin plate.

Dec 7, 2007

No colonies...

Brief Conclusions: No luck this time.

4.3 Cloning ChIP DNA: try 2

Wed Dec 12, 2007

I'm going to try both cutting and not cutting the chromatin this time; I'm also going to try both the 1.5 day samples and the 2.5 day samples. For the 1.5 day, I used samples 2 (lrp) and 9 (pdhR) from section 2.19 on page 169. For the 2.5 day, I used samples 4 (lrp) and 6 (pdhR). As in those sections, I used 16 μl sheared chromatin with 64 μl of dilution buffer for the 1.5 protocol and 20.5 μl sheared chromatin with 59.5 μl dilution buffer for the 2.5 day protocol. I used 2.7 μl antibody (3.3 μg) and 100 μl beads. The only alteration to the protocol was a cutter-wash step in 100 μl of cutter mix after the two TE washes.

I ran the digestions as in the previous section. The digestions were for 15 min at 37C. In addition, for each sample (i.e. 2, 9, 4, 6) I also included a replicate where I did *not* digest the chromatin (I still incubated at 37C for 15 minutes, but without the enzyme).

Thur Dec 13, 2007

I screwed up and forgot to reverse the crosslinks overnite, so I reversed them during the day today for 10 hr and 50 min at 65C followed bh 1 hr at 55C after the addition of proteinase K. I cleaned up the 16 samples with a Qiagen PCR purification kit and eluted into 30 μl of EB buffer.

Fri Dec 14, 2007

I used the Qubit hsDNA kit and 10 μl of each sample to quantify the DNA yield of each ChIP reaction:

sample	**yield (ng)**
1.5 day, lrp 2, antibody, cut	0.588
1.5 day, lrp 2, no antibody, cut	<0.3
1.5 day, lrp 2, antibody, uncut	1.2
1.5 day, lrp 2, no antibody, uncut	<0.3
1.5 day, pdhR 9, antibody, cut	3.216
1.5 day, pdhR 9, no antibody, cut	<0.3
1.5 day, pdhR 9, antibody, uncut	2.832
1.5 day, pdhR 9, no antibody, uncut	<0.3
2.5 day, lrp 4, antibody, cut	2.166
2.5 day, lrp 4, no antibody, cut	<0.3
2.5 day, lrp 4, antibody, uncut	3.99
2.5 day, lrp 4, no antibody, uncut	0.498
2.5 day, pdhR 6, antibody, cut	0.894
2.5 day, pdhR 6, no antibody, cut	0.306
2.5 day, pdhR 6, antibody, uncut	0.336
2.5 day, pdhR 6, no antibody, uncut	<0.3

Brief Conclusions: With no exceptions, the no antibody sample always has less than the antibody sample. The lrp sample always has more DNA when the chromatin is not cut, which makes sense. THe pdhR samples are always the opposite, which doesn't make sense. The yields seem to vary pretty drastically across all of these samples. It's hard to say if this is real, or if I'm just at the edge of the sensitivity of the Qubit dye, so the measurements are noisy. If I used 20 μl of the ChIP sample, I should be able to get a slightly more robust concentration estimate.

I ran out of time to clone this DNA, because my thesis was on Dec 17. I want to try this again with freshly prepared chromatin. I want to quantify them with 10 μl , try qPCR with another 10 μl , and try to clone with the remaining 10 μl . I think I'll switch to using only one restriction enzyme for now to simplify things.

Chapter 5

Experimental Operon Determination

THIS CHAPTER/PROJECT IS DEAD (or at least not-active)

Please see the last paragraph of this chapter for conclusions that might still be interesting regarding what I found in this chapter. It seems that a huge chunk of the bacterial genome is transcribed – even regions that you wouldn't expect. However, I'm leaving this project aside to focus on the PET sequencing project that will hopefully provide similar/better information for less work, less money, and in a more generalized framework. If I pick this project up again, I definitely need to try doing qPCR to these regions.

Tue Feb 14 17:25:50 EST 2006

I'm not interested in operon predictions, known operons, or complex operons. However, as I explained in my Feb 2, 2006 labmeeting talk, it is important to know all of the operons in a genome if you really want to verify predicted regulatory interactions using ChIP. You have to know the most likely binding site of the TF (which is in front of the first gene in the operon that the TF is predicted to regulate) to design a primer set that quantifies the enrichment of your TF for your target. It was for this reason that I initially became interested in developing a high-throughput method of operon determination.

The method I planned to use is simple and the low-throughput version is not-at-all novel. I can't find too many recent papers describing it, but I'd guess it's the standard way operons were found in the past. The strategy is shown visually in Figure 5.1. The idea is this: You have 3 genes you think might be in an operon together. To test this, design 5 sets of primers and make sure the primers work and are specific by testing them on genomic DNA. Using cDNA produced from an mRNA lysate of a particular condition or a mix of several conditions, run a PCR reaction for each of the primer pairs. Amplification of the gene itself (Fig 5.1 orange arrows) is a necessary precursor to determine operons (Fig 5.1 purple arrows). Successful amplification of of the genes lets you know if the gene is present in the cDNA. If it isn't, you can learn nothing about this particular genes operon from this particular cDNA sample. If two consecutive genes are present (two adjacent orange arrow pairs are amplified) and you can amplify from one gene to the next (the purple arrows), then the genes must be transcribed as one unit and are in an operon together. By tiling this procedure along the genome you need 4 primers per gene * 4000 genes = 16000 primers.

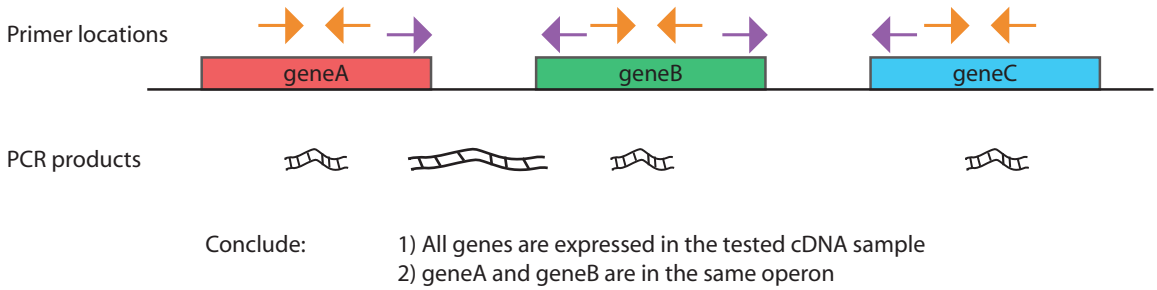


Figure 5.1:

16000 primers is 8000 PCR reactions and 8000 gels. However, by multiplexing the PCR reactions (5-10 per tube) and running microfluidic gels using something like the Agilent 5100 ALP, it should be possible to cut this down to around 1000-1500 PCR reactions. With a mediocre PCR machine and a robot plus the 5100 ALP (or LabChip 90), all the operons for one species could be checked in 24-48 hours. Specifically for the ALP, the cost is $130,000, plus $400 for reagents for 3000 samples, plus $900 for the microfluidics chips to run 6000 samples. Assuming the multiplexing could bring the number of reactions to 1000, the cost per operon validation is $288 – less than a microarray. This means it wouldn't be too hard to run a few studies in different conditions to test for complex operons. There is also an initial fee for the primers needed to do all the PCR reactions: 16000 primers would be $20-40K. The machine can handle about 340,000 samples a year and has a 10% downtime (sale rep says it is a new machine that needs to be repaired a fair amount), still that's 170 potential full operon checks a year.

A smart algorithm could pick primers that would potentially work in multiple related species. Also, the size ranges of the PCR amplification would need to be chosen so that they are different enough

that when each reaction is run on an e-gel all 5-10 bands (for the amplified PCR reactions) are easily distinguishable. The e-gel chip works for 25-1000bp and at least in their literature easily distinguishes stuff more than 7bp difference in length. In practice we'd probably try for at least 50bp between the fragment lengths. To increase the chance of having more genes present and therefore more operons determined, it might be necessary to pool mRNA from many different conditions. This would destroy any complex operon information.

The more I think about this method the more I like it. One thing that it could address better than any previous study is: *what fraction of an organism's genes are present in the cell at any given time.* Microarrays kinda give you this information, but while my experience with Affy chips leads me to believe the are fairly precise for *E. coli*, I have no idea how sensitive they are. PCR is certainly more sensitive, and comparing just the gene amplifications (orange arrows) to the Affymetrix call values (Present, Absence, Marginal) would be interesting by itself. In addition, once you have all these primers you can do some really nice spike-in studies to really test the range of microarray quantitation and how cross-hybridization affects accuracy (see my prospectus notes for details).

This method will not work if a complex operon is transcribed in multiple forms in the same condition. For this method to work it is *VERY IMPORTANT* that the mRNA prep be completely clear of genomic DNA. Otherwise it'll be a real pain-in-the-ass to set some theshold to chop things off that you think are amplified just because of the contamination. I think it would be wise to use the new Qiagen preps that are designed to reduce genomic DNA PLUS a DNAse digestion just to be safe. To check for this, it would probably be useful to run a few PCR reactions on the cleaned-up mRNA to make sure nothing amplifies.

5.1 Initial Tests

I don't think the PI is too fond of this idea. My guess is he thinks it is expensive for what you get, especially since he's not the Broad Institute with deep pockets. However, some of the network inference predictions to unknown targets have a known target right next too them. The simplest explanation for the interaction is that the two genes are in an operon together, which is why the inference algorithm says they are both regulated by the same TF. To validate the new target, you don't need to do a complicated ChIP experiment, all you need is to do a two-gene version of what is shown in Figure 5.1.

I ordered primers to test this for genes flgK and flgL. flgK is a known target of the flagella regulators, flgL is not but it is right next to flgK. Now that I'm typing this, I realize I should order primers for a few other genes just to make sure this works for known genes.

5.1.1 flgK, flgL operon primer tests

Mon Mar 20 12:32:21 EST 2006

I'm going to test the primers by simply doing a PCR of genomic DNA. I'm also going to try doing a completely unintelligent multiplex by dumping all the primers into one reaction assuming no nasty interactions are going to happen. In practice I don't plan on multiplexing the same operon in the same tube, since there might be steric problems with amplifying some many regions that are very close to each other.

I'm running a 25μl PCR with the following reaction:

PCR Reaction composition	
H_2O	8.5 μl
Qiagen Master Mix	10 μl
Forward and reverse primer	1.5 μl
final primer concentration	150 pM
template DNA	100 ng

Thermal cycler conditions		
Initial denaturation	5 min	95°C
2-Step cycling		
Denature:	30 sec	95°C
Anneal/Extend:	45 sec	60°C
Number of Cycles:	28	
Final Extention:	7 min	72°C

The predicted lengths of the amplified regions of each primer pair are:

ID	forward primer	reverse primer	length (bp)
A	flgK F	flgK R	161
B	flgL F	flgL R	220
C	flgK-L F	flgK-L R	307

Eight combinations will be tested: blank, A, B, C, AB, AC, BC, ABC.

Unfortunately, after the PCR finished the ABC reaction tube was empty. I think these PCR tube strips I have are too thin and it must have had a little hole in the tube and the stuff inside leaked out. There was just a tiny amount. The other tubes on the strip looked fine.

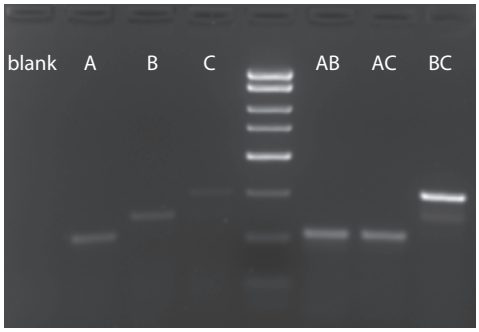


Figure 5.2: 80 ml, 2% agarose gel with 0.5 ul of 1% ethidium bromide run for 40 min at 100 volts. 10 μl of Fisher BioReagents exACTGene Low Range DNA Ladder, with bands:weights(ng) of 2000:105, 1500:87, 1000:68, 750:59, 500:94, 300:27, 150:34, 50:25 was used. Product sizes for the 3 amplifications should be: A=flgK=161bp, B=flgL=220bp, C=307bp. 10μl of the PCR product was run on the gel.

Brief Conclusions: The results certainly were not too bad for a first try. All the single-plex reactions worked and were the correct length and easily differentiable. The multiplex (as explained above this wasn't an ideal multiplex situation) reactions didn't fair as well. But one of the three tested worked. This is only a 2-plex reaction. Unfortunately the 3-plex reaction evaporated. I think another attempt and at the same time a run on cDNA should be attempted. This would be much better if I had a few operons to test so I didn't have to multiplex reactions that fight for the same template DNA initially. I'm not sure what the effect of increasing the genomic DNA and/or descreasing the primer concentration would have. It'd be more appropriate to get two known operons to add in the mix with this unknown one before delving into improving the protocol.

5.1.2 PCR tips

here are some PCR tips I pulled off of this excellent website.

- initial denature doesn't matter. I will reduce to 2 min or remove completely (i.e., start at cycle one with 30-60 sec denature).
- long denature (in cycle) breaks down Taq. only need 30-60 secs at 94C
- anneal at 54C is better for multiplex
- extension at 65C better for multiplex (than the standard 72C)
- extension time increase to 2 min helps multiplex (what about 1 min per kb of multiplex DNA [summing all lengths?])
- template DNA (250-125ng)
- optimize template conc.
- anneal 30-45 secs (lower prevents non-specific amplification)
- primer concentration 60-200

5.2 Moving to a more realistic scale

Tue May 2 16:08:15 EDT 2006

I have to work on a large scale (i.e. $>$ 24 PCR reactions) otherwise optimizing the multiplex PCR to follow size, location, and primer-dimer constraints is very difficult. I'm going to focus on small contiguous chunks of genomes. The first chunk of 19 genes is in Figure 5.3. Working in little chunks is helpful because it makes the primer sets somewhat self-contained (i.e. like a programming function is makes the project seem smaller, stresses the big picture, and localizes errors). Each chunk will occupy a 96-well plate. I haven't yet decided whether the multiplex set will come from combining chunks (to minimize steric constraints) or within a chunk. Making cDNA with random hexamers removes all but the most extreme steric problems (i.e. two amplifications that overlap), but those are trivial to prevent. The only time there might be a problem is with running the positive control on genomic DNA. Last, it is relatively easy to find syntenic regions of 19 genes across multiple organisms allowing the generation of primer sets that work on multiple organisms. Longer stretches run into problems that will have to be dealt with on a per species basis.

Description of Genome Chunk1 Tue May 2 16:08:15 EDT 2006 Genome Chunk 1 contains 19 genes: lpxL yceA yceI yceJ yceO solA yceP dinI pyrC yceB grxB mdtH rimJ yceH mviM mviN flgN flgM flgA

In Figure 5.3, known operons are designated by a gray rectangle background.

Pros:

- few operons of size 1; cool to see if they are incorrectly annotated

- many different functions that we know how to stimulate based on our expression array compendium (particularly the flagellar and dna damage genes)
- the dna damage genes were predicted targets of lexA in the CLR work

Cons:

- only one *known* operon with more than one gene in it flgAMN
- few larger intergenic regions that might contain an unannotated gene (though there are a few little spots)

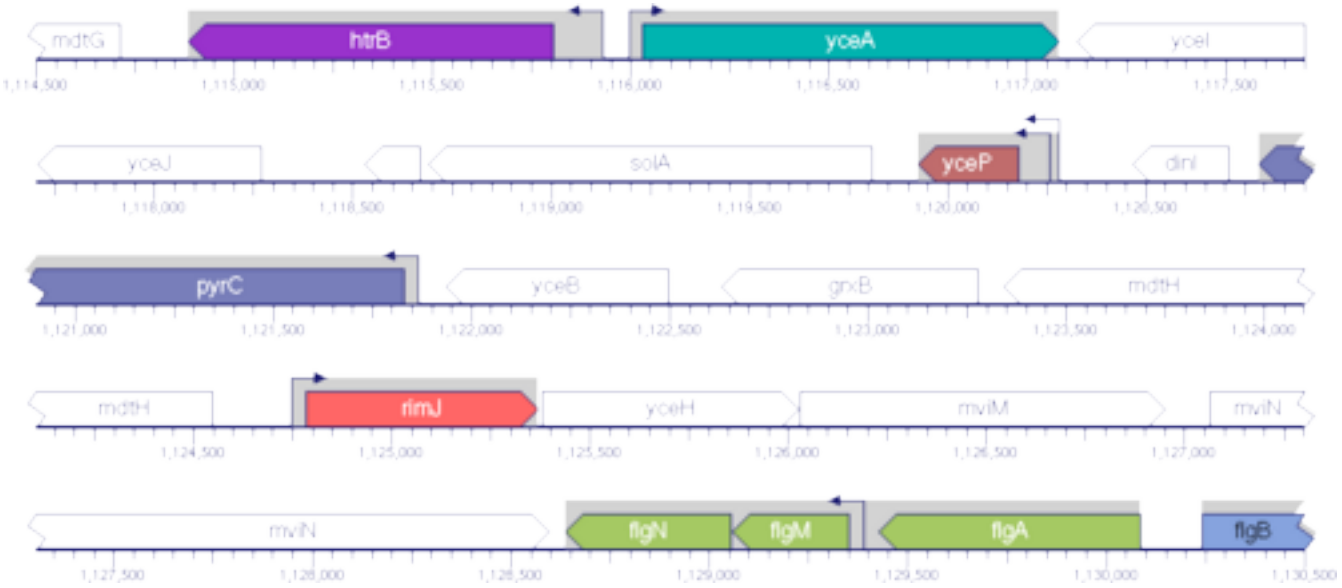


Figure 5.3: Genes whose operon will be tested in the initial pilot study

5.2.1 Getting started

Sun Jun 4 17:27:43 EDT 2006

It was difficult to design multiplex primers using only one chunk in a 96-well plate. So I added a second chunk consisting of the following 11 genes: ydgD, mdtI, mdtJ, ydg, pntB, pntA, ydgH, ydgI, folM, ydgC, and rstA. In total the first plate of operon primers contains 30 genes and 28 gene-spans (the oligos amplifying from the end of one gene to the beginning of the next).

The plates were filled with each columning containing primers that are entirely compatible with each other (based on computational estimates). So each plate can potentially be plexed up to 8 reactions. The rules in designing the primers were:

1. primers must work across W3110 (lab strain), EDL933 (O157:H7 pathogen), MG1655 (most common lab strain), and Sfl2457T (shigella very common 3rd world pathogen)
2. primer size: 23+/-3

3. melting temp 61+/-3
4. GC content 50%+/-10
5. remove first and last 20 base pairs (in case the ends of the genes were incorrectly annotated or unevenly transcribed in the cDNA reaction; making sure the primers work across four species also helps be certain that all the gene boundaries are good.)
6. no self complementation or primer dimer formation across the pairs
7. size between 75 and 650 bp

The rules for multiplexing were:

1. 8-plex to fill up an entire column of a 96-well plate
2. lengths of products must be at least 10% different so they can be differentiated on an agarose gel
3. no primer dimers are allowed between any of the primers in the plexed reaction

The multiplex rules were satified by a greedy search for a set of primers that fit the rules. The search often does not converge to an agreeable solution (i.e. all primer sets are 8-plex), but the search is restarted until an agreeable solution is found (typically 5-30 tries). The software used for this was: primer3 for designing the pairs, ntdpal (basic alignment software that comes with primer3) for checking multiplex primer-dimer formation, and a couple perl scripts to glue the pipeline together.

5.2.2 Preping genomic DNA

Mon Jun 5 17:58:52 EDT 2006

I attempted to prep genomic DNA for use in the PCR experiments with *very little yield*. Up-to the lysis step worked very well. They lysate was clear after 1 hr. Previously the solution had been very viscous even after the prep was finished. I tried to clean it up a little by doing an initial ethanol precipitation followed by incubation with RNAse cocktail to the remove RNA yielding cleaner genomic, but the solution was so goopy that I had a hell of a time getting rid of it without out getting rid of my DNA.

Two preps were done yielding:

Sample	DNA (ng/ul)	260/280	260/230
1	2.2	2.44	0.4
2	36.0	2.29	1.51

Brief Conclusions: The yield was pathetic and handling the DNA in the goopy lysate was a real pain-in-the-ass. Tomorrow I'm going to try again but starting with 2 ml of culture (the original protocol) and going straight to a phenol:chloroform using gel-phase-lock, followed by addition of RNAse Cocktail, followed by a second phenol:chloroform extraction with a gel-phase-lock tube. Maybe if I want to increase the yield with the larger volume, I need to also increase the amount of TE I'm working with. I can double the concentrations and still fit it in a gel-phase-lock 2 ml tube. I just would have to use isopropanol precipitation. I plan on using isopropanol tomorrow anyways (that's what the original protocol suggests).

Brief Conclusions: Tue Jun 6 17:01:03 EDT 2006

Just an update. I think a large part of the problem is that it takes *forever* for the genomic DNA to go back into solution. I rechecked today and there is more DNA in the tube not a huge amount but significantly more than before. Some of the genomic DNA can be seen in Figure 3.8 and again in Figure 5.4. The DNA was much too long for the type and voltage of the gel.

The yields are now:

Sample	DNA (ng/ul)	260/280	260/230
1	28.2	1.57	0.84
2	40.0	1.71	1.08

I think the way to go might be to shear it by sonication after the lysis is complete. That would more represent the cDNA that I'm actually going to be using anyways.

5.2.3 Preping genomic DNA: 2nd try

Tue Jun 6 17:04:17 EDT 2006

Tried 3 samples. One without RNA digestion (sample 1). Two with RNA digestion in between the two phenol:chloroform extractions (samples 2 and 3). The third sample used double of everything (4 ml of culture, etc)

The yields as of Wed Jun 7 16:51:09 EDT 2006 (about 18 hours for resuspension) are:

Sample	DNA (ng/ul)	260/280	260/230	total yield (ug)
1	327.3	2.14	2.32	167.7
2	82.6	1.76	1.36	41.3
3	117.4	1.78	1.37	58.7

All three samples were run on a 1% agarose gel (see Figure 5.4).

Brief Conclusions: I'm pretty confident the genomic prep is working. I think working with genomic DNA sucks because it is goopy, slimy, sticky, hard to pipette, and hard to resuspend. HOWEVER, it looks like the prep is working. For an easier time, I would shear the DNA right after the lysis step. The RNA digestion also seems to be important.

5.2.4 Preping genomic DNA: with shearing

June 12, 2006

I ran the genomic DNA prep protocol but this time I sheared the lysate by sonication after the lysing step. The hope is that the shearing will make the prep easier and the resulting DNA should be more similar to cDNA created by random hexamers (i.e. shorter fragments not continous long pieces). Samples 1-3 were sheared at 20% power for 30 seconds, which resulted in a foamy mess. Samples 4-6 were sheared at 10% for 30 seconds, which worked much better (no foaming).

The yields as of Jun 13 are (samples in 300 μl total volume):

All 6 samples were run on an agarose gel (see Figure 5.5).

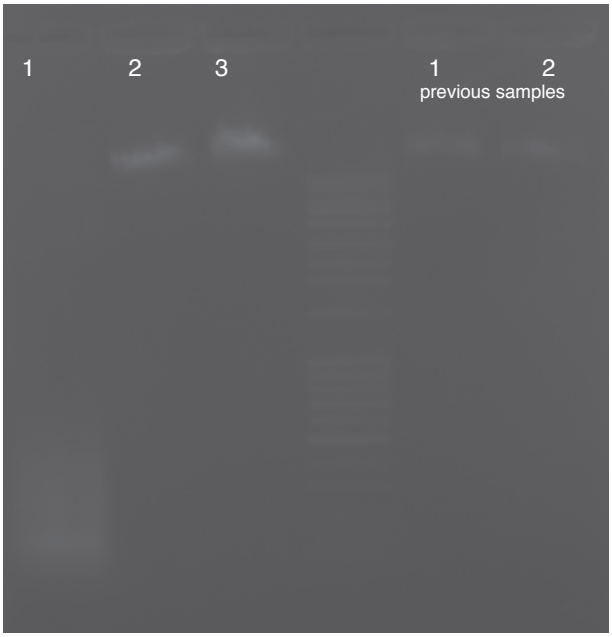


Figure 5.4: Genomic DNA on a 1% agarose gel. The genomic DNA is too long and doesn't run properly. Samples 1 and 2 on the right are from the previous attempt at making genomic DNA. Sample 1 on the left did not have an RNAse digestion step (as can be seen by the large amount of RNA)

Sample	DNA (ng/ul)	260/280	260/230	total yield (ug)
1	186.7	1.81	1.57	56
2	284.5	1.82	1.73	85.4
3	246.1	1.84	1.66	73.8
4	237.2	1.84	1.79	71.2
5	246.2	1.78	1.33	73.9
6	152.9	1.75	1.27	45.9

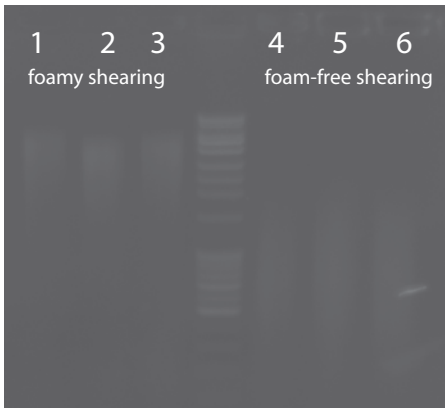


Figure 5.5: Sheared genomic DNA on a 1% agarose gel.

Brief Conclusions: The genomic DNA was easier to handle and sheared into a broad range. Shearing size was better when using 10% power and avoiding the foaming problem. Samples 4-6

will be used as the input DNA to further operon PCR reactions.

5.2.5 Preparing the oligos

The oligos were shipped at 200 uM in two plates (one forward one reverse). I'm going to dilute and mix the forward and reverse primers into a new plate. The mixing is:

- 1 μl of forward primer
- 1 μl of reverse primer
- 198 μl of TE

This gives a final concentration of 1 uM for each primer. For a 5 μl PCR reaction and 150 nM of each primer, I just need 0.75 μl of the mix.

5.2.6 Realistic Scale: first experiments → single-plex

All 58 primers are going to be tested in a single-plex reaction using genomic DNA. I'm going to use sample 2 of the genomic DNA prep on June 6th.

Single-plex first attempt

PCR run on Jun 7, 2008.

The plate is:

Single-plex oligo plate												
-	1	2	3	4	5	6	7	8	9	10	11	12
A	ydgH_106	mdtH_122	solA_77	yceI_191	yceO_75	yceJ_yceO_522	ydgD_mdtI_263	flgM_flgA_261	-	-	-	-
B	pntB_pntA_220	yceO_solA_329	mviN_166	flgN_268	mdtJ_ydgG_623_622	mdtI_174	yceJ_101	ydgI_folM_348	-	-	-	-
C	rimJ_yceH_165	flgM_98	yceH_112	mdtI_mdtJ_216	yceA_yceI_167	ydgH_ydgI_371	mviM_mviN_444	-	-	-	-	-
D	dinI_94	pntA_461	grxB_201	grxB_mdtH_322	lpxL_yceA_466	yceA_424	flgN_flgM_178	-	-	-	-	-
E	yceI_yceJ_121	yceP_dinI_509_512	ydgG_372	pyrC_137	solA_yceP_393	dinI_pyrC_316	mdtH_rimJ_337	-	-	-	-	-
F	flgA_184	pyrC_yceB_363	ydgD_454	lpxL_420	mviN_flgN_341	yceH_mviM_235	folM_389	-	-	-	-	-
G	yceB_grxB_266	mdtJ_170	ydgI_272	mviM_89	rimJ_299	ydgG_pntB_271	yceP_79	-	-	-	-	-
H	ydgC_79	folM_ydgC_195	ydgC_rstA_235	pntB_111	rstA_230	yceB_197	pntA_ydgH_638	-	-	-	-	-

PCR Reaction composition	
H_2O	1.5 μl
NEB PCR Master Mix	5 μl
Forward and reverse primer	(1.5 μl) 150 nM
template DNA	(2 μl) 150 ng

Thermal cycler conditions		
Initial denaturation	30 sec	95°C
3-Step cycling		
Denature:	30 sec	95°C
Anneal:	30 sec	54°C
Extend:	120 sec	65°C
Number of Cycles:	30	
Final Extention:	5 min	70°C

Samples were run on a 2% gel Jun 8, 2006 (see Figure 5.6).

Brief Conclusions: Nothing (not even easy PCR reactions) work right the first time. The lengths of the products in Figure 5.6 don't correspond to the expected product sizes of the tested genes. I'm not sure why. Beyond this problem, the 96-well PCR plate also had some serious evaporation issues on the edges (see Figure 5.7). Overall I think nothing can be concluded from this (except don't use the Costar 96-well plates next time).

Figure 5.6: 300 ml 2% agarose gel run at 4.8 V/cm in TAE. 1.6 μl EtBr. The lengths here don't correspond to the expected lengths. Something is weird here. Missing wells (less than column 8) are due to evaporation in the wells on the edge of the PCR plate.

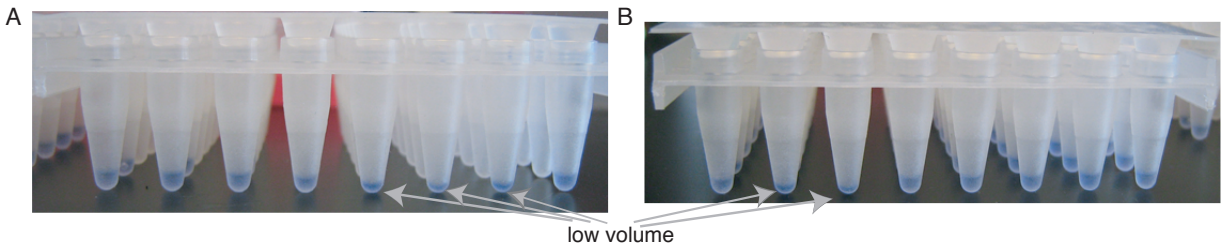


Figure 5.7: The Costar 96-well PCR stuff just doesn't seem to work well on the edges of the plate. I think I need to move to BioRad (the maker of the PCR machine).

Single-plex second attempt

Jun 8, 2006

I switched to 8-well BD Falcon PCR-strips figuring that for sure this would fix the evaporation problems. WRONG!!! Just for the record, BD Falcon PCR strips SUCK!!! How hard is it to make a freaking plastic tube that stays sealed when a heated-lid is pushing down on it? Apparently it is difficult for the folks at BD Falcon. Same evaporation issues I had with the Costar plate occured with the stupid strips (no picture because in my anger I grabbed all the tubes and forcefully threw them away w/o thinking about a picture). I've never had this problem with a PCR strip before.

I lowered the amount of genomic DNA (because I'm running low) and the extension time (to speed things up).

PCR Reaction composition	
H_2O	1.5 μl
NEB PCR Master Mix	5 μl
Forward and reverse primer	(2 μl) 150 nM
template DNA	(1 μl) 75 ng

Thermal cycler conditions		
Initial denaturation	30 sec	95°C
3-Step cycling		
Denature:	30 sec	95°C
Anneal:	30 sec	54°C
Extend:	60 sec	65°C
Number of Cycles:	30	
Final Extention:	5 min	70°C

Valuable Lesson: BD Falcon PCR strips don't keep their caps on tight in a BioRad iCycler.

Single-plex third attempt

Jun 9, 2006

I switched to MP (Molecular BioProducts) PCR strips. And repeated the PCR for all the primers.

PCR Reaction composition	
H_2O	1.5 μl
NEB PCR Master Mix	5 μl
Forward and reverse primer	(2 μl) 150 nM
template DNA	(1 μl) 75 ng

Thermal cycler conditions		
Initial denaturation	30 sec	95°C
3-Step cycling		
Denature:	30 sec	95°C
Anneal:	30 sec	54°C
Extend:	60 sec	65°C
Number of Cycles:	30	
Final Extention:	5 min	70°C

Brief Conclusions: Finally, all the primers amplified the correct size with no failed primer (Figure 5.8). The ladder in the gel is too blurry for the sizing I want to be able to do with the Versadoc software. I think I need to reduce the amount of agarose I use and run at a higher voltage (and maybe try TBE or SB buffer). It's time to multiplex!!!!!

5.2.7 Multiplexing chunks1 and 2

Jun 11, 2006

Of course, with my luck (or perhaps inattention to detail is a better word) I had a bug in my software for choosing multiplex sets. It was supposed to only allow product lengths 10% different from each other in the same well (so they could be differentiated on the gel). In general this was the case EXCEPT, the bug allowed an exception when the product lengths were almost exactly the same length (i.e. around 0% different). This is the worst possible bug since it drastically reduces the options for plexing this thing up.

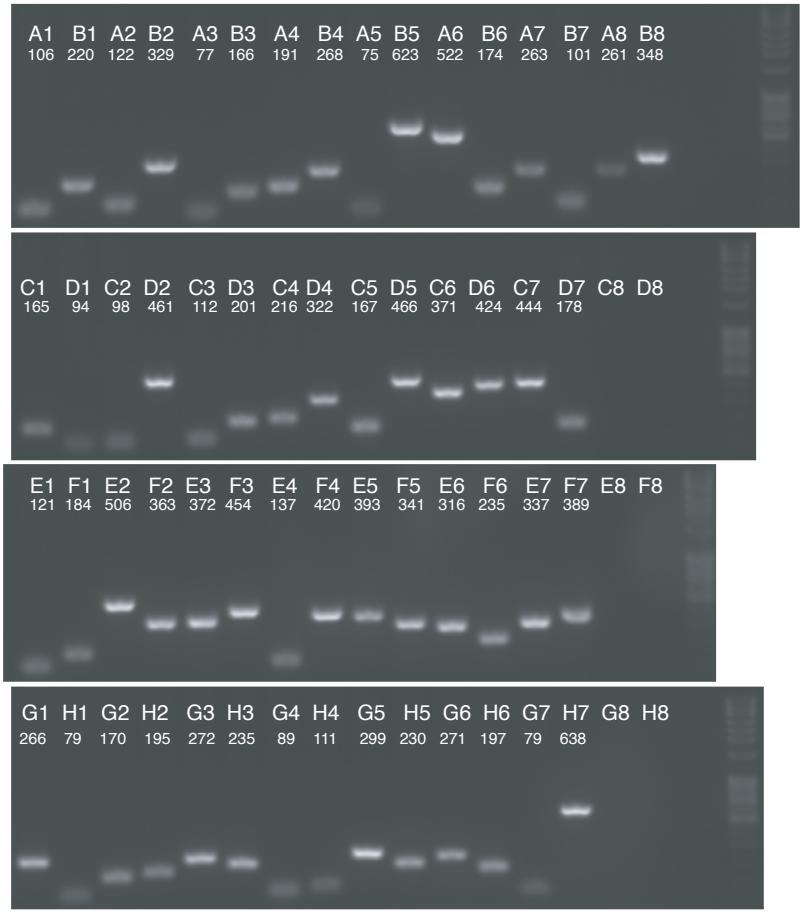


Figure 5.8: 300 ml, 2% agarose gel with 1.6 μl EtBr. The primers all seem to work as designed. Unfortunately the ladder looks like hell.

Still the primers work and I've learned a lot so far and I can still learn from making the best of this situation. Looking by eye I came up with a way to cut the eight rows down to three (one 4-plex and two 2-plexes). The rows are:

α-plex = rows A, B, D, G
β-plex = rows C, H
γ-plex = rows E, F

PCR Reaction composition	
H_2O	1.5 μl
NEB PCR Master Mix	5 μl
Forward and reverse primer	(1.5 μl) 100 nM
template DNA	(2 μl) 150 ng

Thermal cycler conditions		
Initial denaturation	30 sec	95°C
3-Step cycling		
Denature:	30 sec	95°C
Anneal:	30 sec	54°C
Extend:	60 sec	65°C
Number of Cycles:	30	
Final Extention:	5 min	70°C

I ran the remaining 6 μl (including dye) of sample to try and get a nicer gel by running it hotter. I ran them out for a long time to try and resolve the lanes with similar sized fragments (see Figure

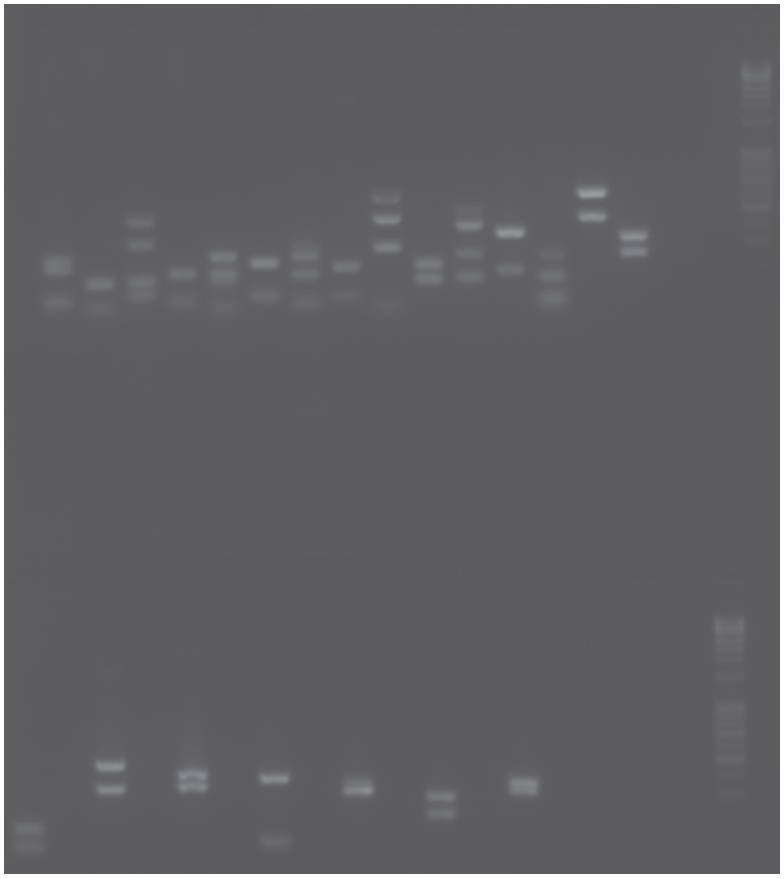


Figure 5.9: 300 ml, 2% agarose gel with 1.6 μl EtBr. Using 1.0 mm comb at 6 V/cm. Here A is α-plex, B is β-plex and C is γ-plex

5.9). I also used the versadoc software to detect the ladder and use it as a standard to estimate the length of each of the bands (see Figure 5.11).

Brief Conclusions: The versadoc software isn't dead-on with the bp estimates (could be my fault if gel is slightly angled or could just be the imprecision in agarose gels). However, it is close enough (+/- 40bp) and the ratios of lengths within a well are good enough to computational know if there are bands of a particular known length or not. I think ALL of the multiplex reactions worked, which is great and all I could wish for. The only problem is that the mixes I picked by hand weren't the best. Columns 1 and 13 (A1 and A7) appear to have only three bands, but I think they are two bands of damn near the same size (A1: 94bp, 106bp; A7: 79bp and 101bp). If I had a clearer gel, I think the A7 should be resolvable (and if you look real close and use your imagination a little you can see them both already). I think running hotter with less agarose and less sample (Figure 5.10) helped out a bit (think slow gel = Figure 5.9. Certainly the move to the 1 mm comb cleans things up. It's only the smaller bands that are smearing. I think 7 μl total volume (including the dye), a 200 ml gel (or 180 ml but that's a pain to measure) will help more. Also need to try TBE and SB buffers and raising the temperature even more. Thankfully the PCR bands are much clearer than the ladder so all I have to do is clean up the thing that's normally the cleanest thing on the gel. Last, if I'm lowering the volume the Fisher dye isn't concentrated

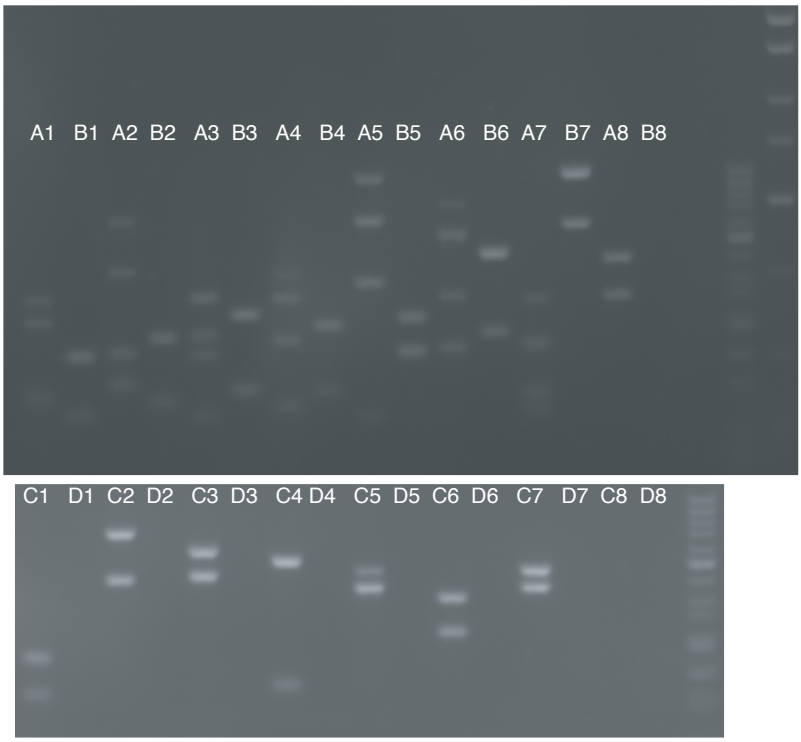


Figure 5.10: 250 ml, 2% agarose gel with 1.6 μl EtBr. Using 1.0 mm comb at 8 V/cm. Here A is α-plex, B is β-plex and C is γ-plex

enough. I'll have to make the dye from DNA by mixing it with buffer. Also, I might try one of the more expensive low-bp agarose gels. But they are almost 2x the cost (making a gel can be almost $8)

Better looking, faster running 96-well multiplex gels

Sat Jun 17, 2006 A 25 μl multiplex PCR was run to allow different gels to be tried on the same DNA.

The set-up of the multiplex plate is:

2-plex oligo plate												
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	ydgH_106	mdtH_122	solA_77	yceI_191	yceO_75	yceJ_yceO_522	ydgD_mdtI_263	flgM_flgA_261	-	-	-	-
	pntB_pntA_220	yceO_solA_329	mviN_166	flgN_268	mdtJ_ydgG_623_622	mdtI_174	yceJ_101	ydgI_folM_348	-	-	-	-
B	rimJ_yceH_165	flgM_98	yceH_112	mdtI_mdtJ_216	yceA_yceI_167	ydgH_ydgI_371	mviM_mviN_444	-	-	-	-	-
	ydgC_79	folM_ydgC_195	ydgC_rstA_235	pntB_111	rstA_230	yceB_197	pntA_ydgH_638	-	-	-	-	-
C	dinI_94	pntA_461	grxB_201	grxB_mdtH_322	lpxL_yceA_466	yceA_424	flgN_flgM_178	-	-	-	-	-
	yceB_grxB_266	mdtJ_170	ydgI_272	mviM_89	rimJ_299	ydgG_pntB_271	yceP_79	-	-	-	-	-
D	yceI_yceJ_121	yceP_dinI_509_512	ydgG_372	pyrC_137	solA_yceP_393	dinI_pyrC_316	mdtH_rimJ_337	-	-	-	-	-
	flgA_184	pyrC_yceB_363	ydgD_454	lpxL_420	mviN_flgN_341	yceH_mviM_235	folM_389	-	-	-	-	-

To make this plate from the single-plex plate you need:

New Row	Old Rows
A	A, B
B	C, H
C	D, G
D	E, F

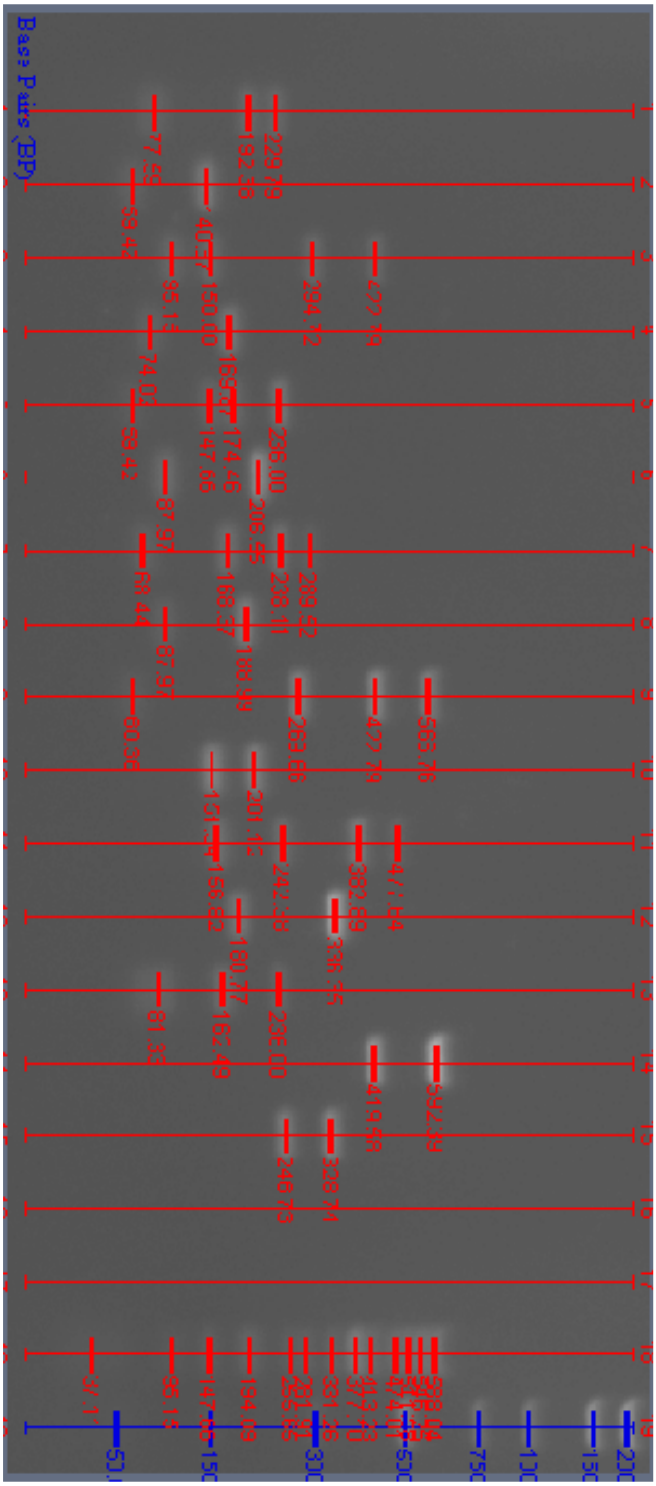


Figure 5.11: Multiplex results annotated by the versadoc.

Mon Jun 19, 2006

The multiplex is working well, but the bands are all fuzzy. The 1 mm comb improved things, but now I'm trying new buffers that are more recommended for shorter DNA and can run at higher

voltages without melting. In addition, the volume is being reduced to make a thinner gel which is clearer. The first gel was run on TBE. The second was run on SB buffer (this is a recently published buffer see B.3.3). Both gels were 180 ml 2% agarose gels run for 40 min at 250 V (10V/cm) with 1.6 μl EtBr.

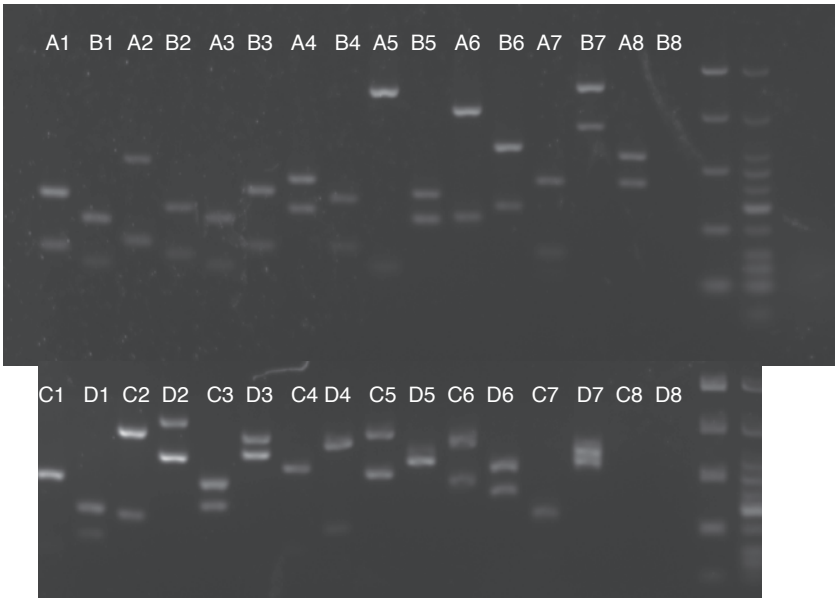


Figure 5.12: Multiplex results on a TBE gel. 2% agarose, 1.6 μl EtBr. 10 V/cm (250 V total). 40 min. 6 μl PCR sample.

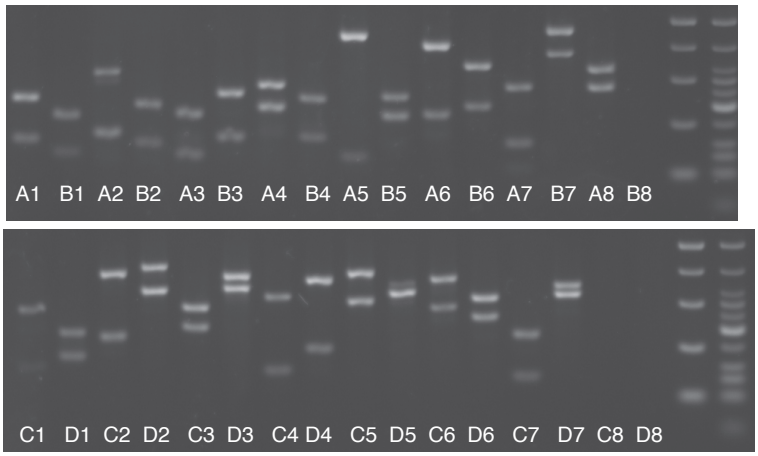


Figure 5.13: Multiplex results on a SB gel. 2% agarose, 1.6 μl EtBr. 10 V/cm (250 V total). 40 min. 6 μl PCR sample.

Brief Conclusions: *The good:* both gels ran in 40 minutes, didn't melt, and were much clearer than the TAE gel (particularly the ladder) (see Figures 5.12 and 5.13. I changed a lot of variables (ran at a higher voltage, different buffer, thinner gel), so it is hard to know which variables increased the clarity of the gel and how much. However, the TBE gel was run near the maximum voltage

for that buffer type while SB should be able to run much hotter (though the power supply we have will only go 50 V higher). The SB gel was also clearer than the TBE gel and SB is cheaper than TBE so that's probably my buffer of choice for now. The software based estimates of DNA length are closer with these clearer gels. *The bad:* The plex in well C1 is barely visible. I might need to increase the amount of sample I put on the gel.

5.3 Moving on to cDNA from a few conditions

Everything seems to be working now on genomic DNA. I can prep sheared genomic with no problems, I can get 2-plex reactions to work on the 60 primer sets I have for the two gene chunks, I have gels that run fast, have good resolution, and use a minimal amount of agarose. Now its time to really see if this thing works! Can this multiplex PCR method work on cDNA to differentiate the operons for these 20-30 genes being tested here???? Finally, a time for results that will add something to knowledge about the cell (even though it is still only a small set of genes).

5.3.1 Initial chosen conditions

Four conditions are being chosen around the two things I know will cause differential expression. 1) +/- norfloxacin will influence the DNA damage operon 2) +/- amino acids will induce the lrp regulated operon.

The four conditions to be tested are:

1. LB
2. LB with 75 ng/ml norfloxacin
3. minimal Davis media with glucose (0.5%)
4. minimal Davis media with glucose (0.5%) casamino acids (0.2%)

I'd prefer minimal media throughout, but I don't want to run to many experiments to start off with and using rich-undefined media like LB provides an easy way to really change the expression of a lot of genes relative to the minimal with glucose so hopefully, I'll really see a lot of genes in their on and off over the course of the 4 experiments. Minimal with casamino acids provides nitrogen and amino acids. LB provides more amino acids, peptides, and vitamins (from the yeast extract). There is also a pH difference between the buffered Davis media and the unbuffered LB media.

5.3.2 Growing cells and RNAprotect

5 ml of cells were grown overnite in LB. The next day 2 ml was washed in Davis media. 250 μl was placed into 25 ml of media (1/100 dilution) of one of the four media from section (5.3.1) (in a 125 ml baffled flask).

Below are the OD measurements (minus background) of the cells. *Emphasized text* is when the sample was taken.

Post-incubation time (hr:min)	Sample1 OD	Sample2 OD	Sample3 OD	Sample 4 OD
2:00	0.29	0.254	0.027	0.083
2:21	*0.412*	*0.42*		
2:44			0.031	0.168
3:35	0.892	0.505	0.036	0.386
3:45				*0.451*
15:31	1.734	1.142	*1.266*	1.646

For samples 1, 2, 4, 2 ml of RNAprotect was added to a 15 ml centrifuge tube and mixed with 1 ml of culture. For sample 3, I mixed 400 μl of sample with 800 μl of RNAprotect. For all samples, I followed the RNAprotect instructions and the samples were placed in the freezer where they are supposely good for 2 weeks (i.e. I can wait two weeks before doing an RNAeasy prep). As the table above shows, sample 3 with glucose only was growing *very* slow. Next time I should use a 1/50 or a 1/25 dilution for this condition or start earlier. As it was, it was getting late and I decided to go with a late stationary sample the next morning. Hopefully, the lrp controlled amino acid genes will still be expressed.

Brief Conclusions: I tracked the LB samples longer because I was afriad the norflox wasn't doing anything. With more time, it was clear it was doing something. As time went on the LB + norflox ended up being passed in OD by the LB only culture. Hopefully, the use of late stationary phase minimal + glucose won't mess anything up.

5.3.3 RNApreps and cDNA

Sun Jul 23 18:38:09 EDT 2006

Been busy with the network inference paper resubmission. But the RNAprotect says it's good for 2 weeks at -20°C. It's only been a week. I used the lysozyme + proteinase K protocol in the RNAprotect manual for the lysis step and then followed the RNAeasy protocol for the rest. I used 1 mg/ml lysozyme in TE and 10 μl of 20 mg/ml Proteinase K per ml of lysis solution. Yields after the RNA easy kit were a little low:

Sample	RNA (ng/ul)	260/280	260/230
1	453.9	2.05	2.34
2	490.8	2.04	2.31
3	316.7	2.07	2.21
4	507.9	2.04	2.00

I want the RNA to be very clean (i.e. free of DNA) so I used the DNA-free kit from Ambion on the RNA samples after the RNApreps were finished (I used 35 min at 37°C incubation instead of 30 min). Yields dropped quite a bit more after using the DNA-free kit, either because you always lose stuff when you switch tubes or because there was a fair amount of DNA probably both. The yields are still plenty high to get a good amount of cDNA:

Sample	RNA (ng/ul)	260/280	260/230	Amount (μl) for 500 ng
1	322.6	2.08	2.15	1.55
2	138.0	1.97	1.64	3.62
3	107.1	1.97	1.55	4.67
4	161.2	1.99	1.52	3.10

500 ng of RNA was used in a Superscript II cDNA reaction according to the Invitrogen instructions. 100 ng of random hexamers and 10 mM dNTP were used. The RNA was placed in -80C and the cDNA was placed in -20C until the PCR reactions are done (probably tomorrow).

Brief Conclusions: Everything looks good as far as the preps go. I still need to run a little on a gel and have a look tomorrow.

Brief Update *Mon Jul 24 13:45:15 EDT 2006*: I noticed (a little late unfortunately), that Ambion also sells a newer product called *TURBO DNA-free* that is supposed to eliminate much more genomic DNA. If the DNA-free kit I tried isn't sufficient I'll should try that one.

Wed Jul 26, 2006

Finally got around to running those RNA samples on a gel. Ran 5 μl of each (so the total amount of RNA was different for each (Figure 5.14).

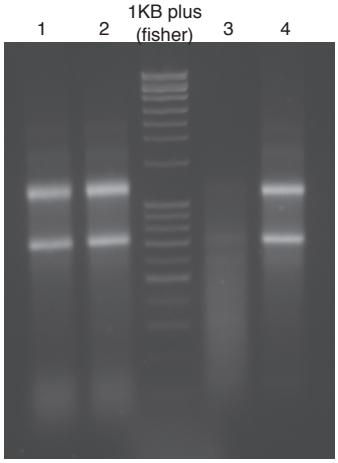


Figure 5.14: 1% agarose, 0.5 μl EtBr. (100 V total). 60 min. 5 μl RNA sample

Brief Conclusions: Things are looking a little sparse in Sample 3 (Figure 5.14. That is the stationary phase culture, so perhaps it's not too surprising. Even the rRNA bands are week though.

5.3.4 2-plex PCR on cDNA sample 3 (norflox)

July 24, 2006

I'm going to use sample 2 (LB + norflox) in a first test of the multiplex with cDNA. Genomic DNA will be used as a positive control. Sample 2 RNA only will be used as a negative control (to make sure the genomic contamination left in my prep is sub-detection by PCR and agarose gel-electrophoresis).

I'm using the 2-plex oligoplate (section 5.2.7 page 220) for the + control samples and sample 2. For - control samples, I'm using just rows C and D from the single-plex plate (section 5.2.6 page 215).

The positive control params (using genomic 4 from page 214) were:

PCR Reaction composition	
H_2O	3.7 μl
NEB PCR Master Mix	5 μl
Forward and reverse primer	(1 μl) 75 nM
template DNA	(0.3 μl) 75 ng

Thermal cycler conditions		
Initial denaturation	30 sec	95°C
3-Step cycling		
Denature:	30 sec	95°C
Anneal:	30 sec	54°C
Extend:	60 sec	65°C
Number of Cycles:	30	
Final Extention:	5 min	70°C

For Sample 2, I added 4 μl of cDNA to the entire master mix (this is 4 μl of the 20 μl total cDNA reaction; into a master mix for 68 samples). For the RNA - control, I added 1 μl of RNA to the entire master mix (for 17 samples), which is still more RNA than is present in the cDNA mixture.

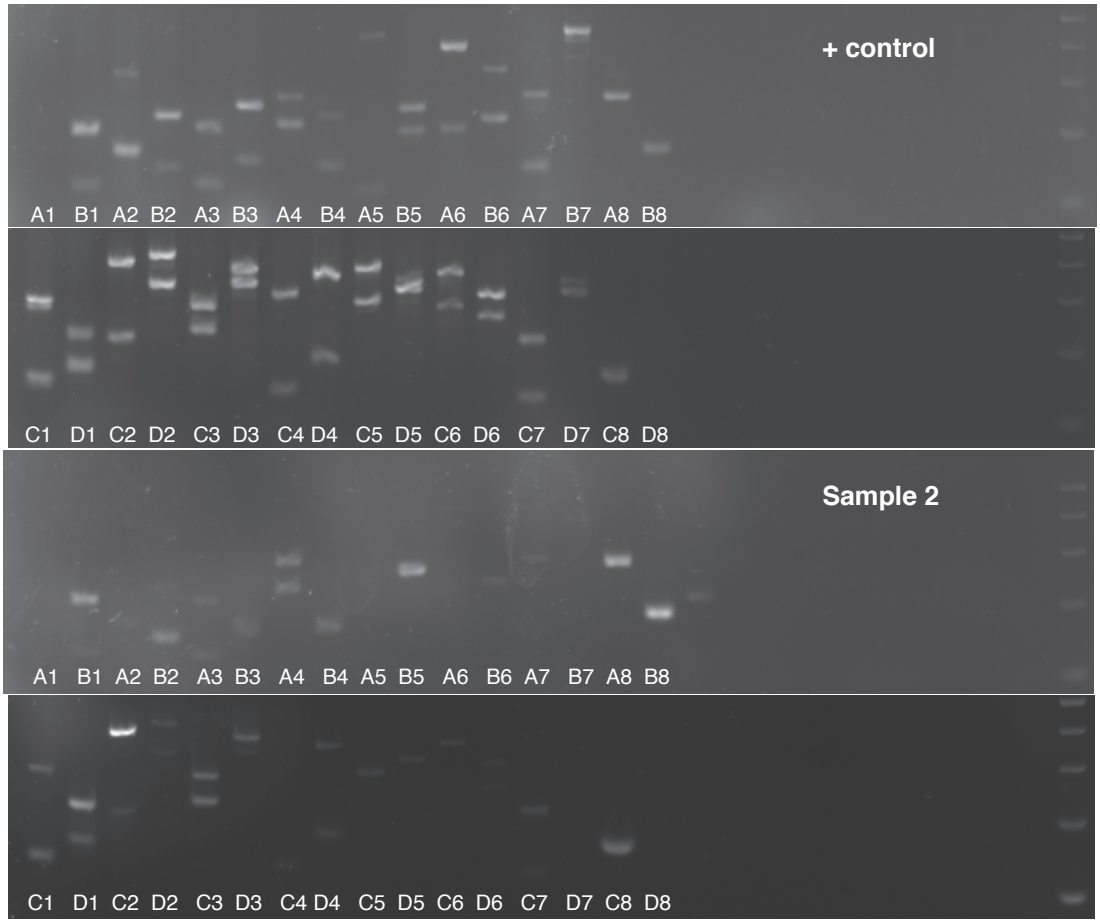


Figure 5.15: 2% agarose, 1.6 μl EtBr. SB buffer (250 V total). 45 min. 7 μl RNA sample; top two panes are + control on genomic DNA

Brief Conclusions: It looks like there might be a little evaporation of the primers in my oligo plate that I keep in the fridge. I much prefer to store my working stock in the fridge, but perhaps in the future I should be more attentive to use it up faster. The gels, though not the most beautiful, do portray what I'd expect. Almost all the + control wells worked (those that didn't had evaporation

issues) (Figure 5.15). Then the PCR reactions on the cDNA are more sparse, showing that not all genes (and operons!) are active (or connected). I will save all my gels to a file in the future, but as I'm going to be running more and more I'll probably not put them all in this document otherwise it'll take too much time and the real analysis in the near future is not going to be done looking at these gels (otherwise I'll never be able to scale up) but rather in the tab-delimited text file I export from the BioRad Versadoc software that gives me the estimated length of each band and the intensity.

Brief Update *Sat Jul 29 20:55:11 EDT 2006*: The aluminum cover is definitely NOT sufficient to prevent evaporation; well A1 was completely dry and the other wells are much lower than they should be given the amount of primer I've used so far. I bought some polypropelyene mats (like how the primers from IDT come) that are supposed to work much better.

Valuable Lesson: When storing reagents in a 96-well plate at RT or in the fridgerator, use a siliconized or a polypropelene mat to prevent having your sample evaporate.

5.3.5 2-plex PCR on cDNA all samples

July 25, 2006

I ran 20 μl of all primers on on all 4 samples. 8 reactions were run on each of the RNA samples.

July 26, 2006

I ran an two SB gel with 7 μl of each sample 5.16. The gel was prestained with EtBr.

The raw data is: test

Brief Conclusions: talk about results and dif between sybr and etbr, that I need to repeat fixing the dried out plate problem and with replication. 4 new conditions

5.4 cDNA from 4 conditions with replication

5.4.1 Growth and RNAProtect

Fri Aug 4, 2006

Due to problems with consistency across the SYBR gold and EtBr gels and what looked possible genomic DNA contamination in my RNA preps (only detectable with SYBR gold), I'm making the following modifications: (1) I'm running 3 replicates of each condition so I can better judge consistancy (2) I'm using the DNA-free TURBO kit which is supposed to be much much better than the previous DNA-free (not turbo) kit that I used to get rid of genomic DNA in my RNA prep. Regarding point (2), I also used the lengthier protocol where you add 1 μl of DNAase, incubate 30 min, add another 1 μl of DNAse, incubate another 30 min. This is 2x the DNAse I used before.

I modified the conditions slightly to hopefully give me a little more diversity in expression that the previous four:

1. LB broth with 100 ng/μl norfloxacin (this is up 25 ng/μl from before); log phase
2. Davis minimal media with 0.5% glucose (as in previous experiment BUT in this experiment I sampled them in log phase)
3. EC broth; log phase
4. EC broth and 100 ng/μl norfloxacin; stationary phase (they had kinda a weird stationary phase at a very low OD; I think 100 ng/μl norfloxacin was pretty harsh so they could only reach around OD 0.2 (see table below and Figure 6.18)

An overnite culture was washed in Davis minimal media and used as a starting culture into the 25 ml media volume. Cells were grown in 250 ml baffled flasks (I wanted to use 125 ml baffled flasks but all of them were dirty). For the minimal media cultures I used a 1/25 dilution from the washed overnite. For the other conditions, I used 1/100.

Growth with time is shown below. Samples taken are shown in *italics*.

***E. coli* growth for 4 conditions 3 replicates**													
Time	**min**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
13:50	70	0.063	0.056	0.07	0.08	0.076	0.08	-	-	-	-	-	-
14:20	100	-	-	-	-	-	-	0.095	0.064	0.099	0.109	0.099	0.071
14:50	130	0.191	0.188	0.226	0.09	0.07	0.084	-	-	-	-	-	-
15:15	155	-	-	-	-	-	-	0.212	0.157	0.235	0.177	0.186	0.202
15:40	180	0.385	-	0.402	0.092	0.078	0.085	-	-	-	-	-	-
15:51	191	*0.418*	*0.403*	*0.467*	-	-	-	-	-	-	-	-	-
16:00	200	-	-	-	-	-	-	0.412	0.323	0.457	0.234	0.26	0.237
16:15	215	-	-	-	-	-	-	*0.533*	*0.381*	*0.523*	-	-	-
16:50	250	-	-	-	0.117	0.114	0.114	-	-	-	0.234	0.239	0.312
17:30	290	0.405	0.31	0.325	-	-	-	-	-	-	0.204	0.222	0.377
17:45	305	-	-	-	-	-	-	-	-	-	*0.195*	*0.191*	*0.398*
17:50	310	-	-	-	-	-	-	0.81	0.868	0.858	-	-	-
18:00	320	-	-	-	0.193	0.193	0.187	-	-	-	-	-	-
18:40	360	-	-	-	0.239	0.271	0.265	-	-	-	-	-	-
19:10	390	0.3	0.213	0.256	-	-	-	1.028	1.03	1.02	0.163	0.159	0.621
19:25	405	-	-	-	0.399	0.4	0.386	-	-	-	-	-	-
19:30	410	-	-	-	-	-	-	1.081	1.072	1.027	0.154	0.145	0.637
19:50	430	-	-	-	*0.42*	*0.473*	*0.461*	-	-	-	-	-	0.75
21:51 (next day)	>800	-	-	-	1.598	1.524	1.335	-	-	-	-	-	1.85

For the EC and LB+nor conditions I added 1 ml of culture to 2 ml RNAprotect. For the Davis condition, I used 1.5 ml of culture and 3 ml of RNAprotect. For the EC+nor sample (which wouldn't reach a high OD), I used 2ml of sample in 4 ml of RNAprotect

Brief Conclusions: So far so good it seems; I'm a little worried that the EC+nor cells are too sickly. I will know more after looking at the RNA on a gel. Also, the EC norflox sample 3 (see Figure 6.18) grew much better than the other two. I only added 2 μl of my stock norfloxacin to reach the 100 ng /μl . I should probably have diluted it a bit. I'm afriad that the difference in growth is likely do to my pipetting error, since it was very hard to get the little pipette into the bottom of my 250 ml flask. THe growth curves aren't the best, since I didn't sample often or long enough, but that really wasn't the point. They're good enough to get the idea that norfloxacin really slows them down at 100 ng/μl and even more so when there are bile salts in the media (for the EC broth).

5.4.2 RNA preps, DNAse digestion and cDNA

Sat Aug 5 21:07:10 EDT 2006

Samples were randomized before RNApreps as:

Condition	Original ID	Randomized ID
LB and 100 ng/μl nor	1	12
LB and 100 ng/μl nor	2	2
LB and 100 ng/μl nor	3	6
Davis and 0.5% glucose	4	5
Davis and 0.5% glucose	5	11
Davis and 0.5% glucose	6	8
EC	7	4
EC	8	1
EC	9	10
EC and 100 ng/μl nor	10	3
EC and 100 ng/μl nor	11	7
EC and 100 ng/μl nor	12	9

This randomization should prevent biases in miniprep, PCR, or gel order from being systematic (as opposed to just random noise). I'll use the randomized order from here on.

For the RNAeasy preps I used 1 mg/ml lyzozyme in TE with 10 μl of Proteinase K (20 mg/ml) for each 100 μl of lysozyme. I had a little problem with the pellets not dissolving easily (they did eventually). Maybe I let the lyzozyme stay with the proteinase K too long before I started (apprx 10 min) and it chopped up the lyzozyme?

I prepped all 12. Then I used Ambion's DNA-free TURBO kit (got a free sample one supposed to work much better than the original DNA-free kit). I used 1 μl TURBO DNAse incubated 30 min. I added additional 1 μl TURBO DNAse and incubated an additional 30 min.

Yields for the 12 samples, post-DNAase digestion are:

RNA prep yields post DNAse digestion						
Sample ID	**ng/μl**	**A260**	**260/280**	**260/230**	**Constant**	**ul in 1 ug**
1	98.2	2.455	2	1.79	40	10.2
2	112.67	2.817	2.04	1.6	40	8.9
3	280.49	7.012	2	1.79	40	3.6
4	675.74	16.894	2.03	2.09	40	1.5
5	362.49	9.062	1.97	1.88	40	2.8
6	451.65	11.291	1.97	1.95	40	2.2
7	222.91	5.573	1.97	1.5	40	4.5
8	370.16	9.254	1.96	1.71	40	2.7
9	532.62	13.315	1.91	1.9	40	1.9
10	424.56	10.614	1.94	1.7	40	2.4
11	287.9	7.198	1.8	1.21	40	3.5
12	373.7	9.343	2	1.83	40	2.7

Add gel of the RNA

I made cDNA with 1 ug of RNA for each sample (final RNA conc in the cDNA is appx 50 ng/μl), this is 2x the amount I used last time, so hopefully this will improve yield a little on the PCRs. I also made an RNA + water negative control sample (the cDNA minus mix) made 30 μl of 50 ng/ul solution. I will used this sample RNA sample for running a negative control PCR.

5.4.3 384-well qPCR

Mon Aug 7 11:55:35 EDT 2006

I'm going to use the new 384-well block on our Icycler PCR machine. I'm using a total volume of 14 μl . Since there are 4 x 8 samples, I can run 12 samples per plate like so:

Layout for 384-well PCR of operon samples																								
-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
A	1	1	1	1	3	3	3	3	5	5	5	5	7	7	7	7	9	9	9	9	11	11	11	11
B	2	2	2	2	4	4	4	4	6	6	6	6	8	8	8	8	10	10	10	10	12	12	12	12
C	1	1	1	1	3	3	3	3	5	5	5	5	7	7	7	7	9	9	9	9	11	11	11	11
D	2	2	2	2	4	4	4	4	6	6	6	6	8	8	8	8	10	10	10	10	12	12	12	12
E	1	1	1	1	3	3	3	3	5	5	5	5	7	7	7	7	9	9	9	9	11	11	11	11
F	2	2	2	2	4	4	4	4	6	6	6	6	8	8	8	8	10	10	10	10	12	12	12	12
G	1	1	1	1	3	3	3	3	5	5	5	5	7	7	7	7	9	9	9	9	11	11	11	11
H	2	2	2	2	4	4	4	4	6	6	6	6	8	8	8	8	10	10	10	10	12	12	12	12
I	1	1	1	1	3	3	3	3	5	5	5	5	7	7	7	7	9	9	9	9	11	11	11	11
J	2	2	2	2	4	4	4	4	6	6	6	6	8	8	8	8	10	10	10	10	12	12	12	12
K	1	1	1	1	3	3	3	3	5	5	5	5	7	7	7	7	9	9	9	9	11	11	11	11
L	2	2	2	2	4	4	4	4	6	6	6	6	8	8	8	8	10	10	10	10	12	12	12	12
M	1	1	1	1	3	3	3	3	5	5	5	5	7	7	7	7	9	9	9	9	11	11	11	11
N	2	2	2	2	4	4	4	4	6	6	6	6	8	8	8	8	10	10	10	10	12	12	12	12
O	1	1	1	1	3	3	3	3	5	5	5	5	7	7	7	7	9	9	9	9	11	11	11	11
P	2	2	2	2	4	4	4	4	6	6	6	6	8	8	8	8	10	10	10	10	12	12	12	12

Brief Conclusions: Thu Aug 24 19:55:43 EDT 2006

I haven't been keeping detailed notes of my results lately. In summary, I've gotten the PCR volume down to 8μl with no evaporation problems (I prepare 4 extra reactions (i.e. for each 32 gene set: one condition; I prepare enough for 36 genes) for pipetting error so I don't run out of master mix. For these reactions I use 0.8 μl primer (100 nM), 4 μl master mix, 3.2 μl H_2O . For each 36 sample set, I use 2.3 μl of cDNA - total.

I ran 10 of the conditions for all 32 genes, each with a negative control RNA only sample (see the gelResults in binary format). There are some problems with genomic contamination, see the binary format data and Figure 5.19.

Things I learned: (1) you can run a SB buffered owl large gel for 40 min at 300 V (12V/cm) for nice separation and relatively fast running time. SB buffer can be used at least 4 times with no noticable effect in gel quality. If SB buffer sits in the rig for more than a couple days it gets a cob-web, white moth-ball kinda substance in it (I always switch buffers when this happens). (2) I seem to be able to amplify most of the genes and a significant part of the operon-spanning fragments (3) EtBr is easier to work with than SYBR the SYBR gold stained stuff shows too much and and it is hard to tell the difference between PCR crap and real signal; any time there really was a real signal I could also see it in the EtBr gel (see Figure 5.19).

Problems I encountered: (1) the PCR machine must not be consistent across the wells. The left half of the gel is always has faint amplification (and therefore faint bands) relative to the right. If I load the samples in the gel in the opposite order the gradient goes the other way, so it seems the problem is not related to the gel itself. (put gel figure and reversed gel figure). I called Bio-Rad and they sent a replacement heat block. I tried again today with the new block and the problem doesn't seem to have been fixed. My only hope is to try the block in the other PCR machine. (2) There was some amplification of the RNA-only control. So the TURBO DNA-free kit was not sufficient to remove all the genomic contamination. (3) I've read a few papers on transcript read through, and it seems like it is pretty frequent for intrinsic terminator sequences (though the papers are based on only a few examples, the almost always find read through). I'm afriad this makes it almost impossible with gel-based methods to determine if the operon spanning region is amplified because the gels are being transcribed as a single unit or because the first gene is reading through its terminator sequence sometimes. The problem might be overcome by qPCR, but then

scaling up is impossible as qPCR master mix would be prohibitively expensive. I plan to try and do paired-end tag (PET) sequencing, hopefully using a highly parallel sequencing method. Now I'm probably going to move this project towards a small study on these 60 genes and intergenic regions using qPCR perhaps with absolute quantitation. One interesting thing would be to use qPCR, microarrays, and sequencing on the same 4 conditions and see how the quantities match up. But for now I'm going to try qPCR. The cloning and preparation of cDNA for sequencing will be reported in its own Chapter. By sequencing PETs you determine both the start and end of a transcript, which if I can get long cDNA (close to full length hopefully) will provide a real nice way to determine operons.

Interesting things: I seem to be able to amplify very often (see binary expression file) a region spanning the 5' regions of two genes that are on opposite strands, here there shouldn't be a transcript read through problem, so it could be any interesting result. Perhaps and interesting direction for this work would be to only design primers for genes on opposite strands that are 5' to 5', and see how often this occurs. I'm going to try and make a few RNA samples that have more negligible amounts of cDNA and test these regions again to make sure this is a *real* thing I'm finding and not some genomic contamination kinda trick. If it's real, I'll probably buy a few primers for some other reginos and try this out in a separate chapter of my lab notebook. Could do a Northern blot to try to determine how long the RNA fragments are that are in these weird regions.

5.5 gDNA free RNA

test

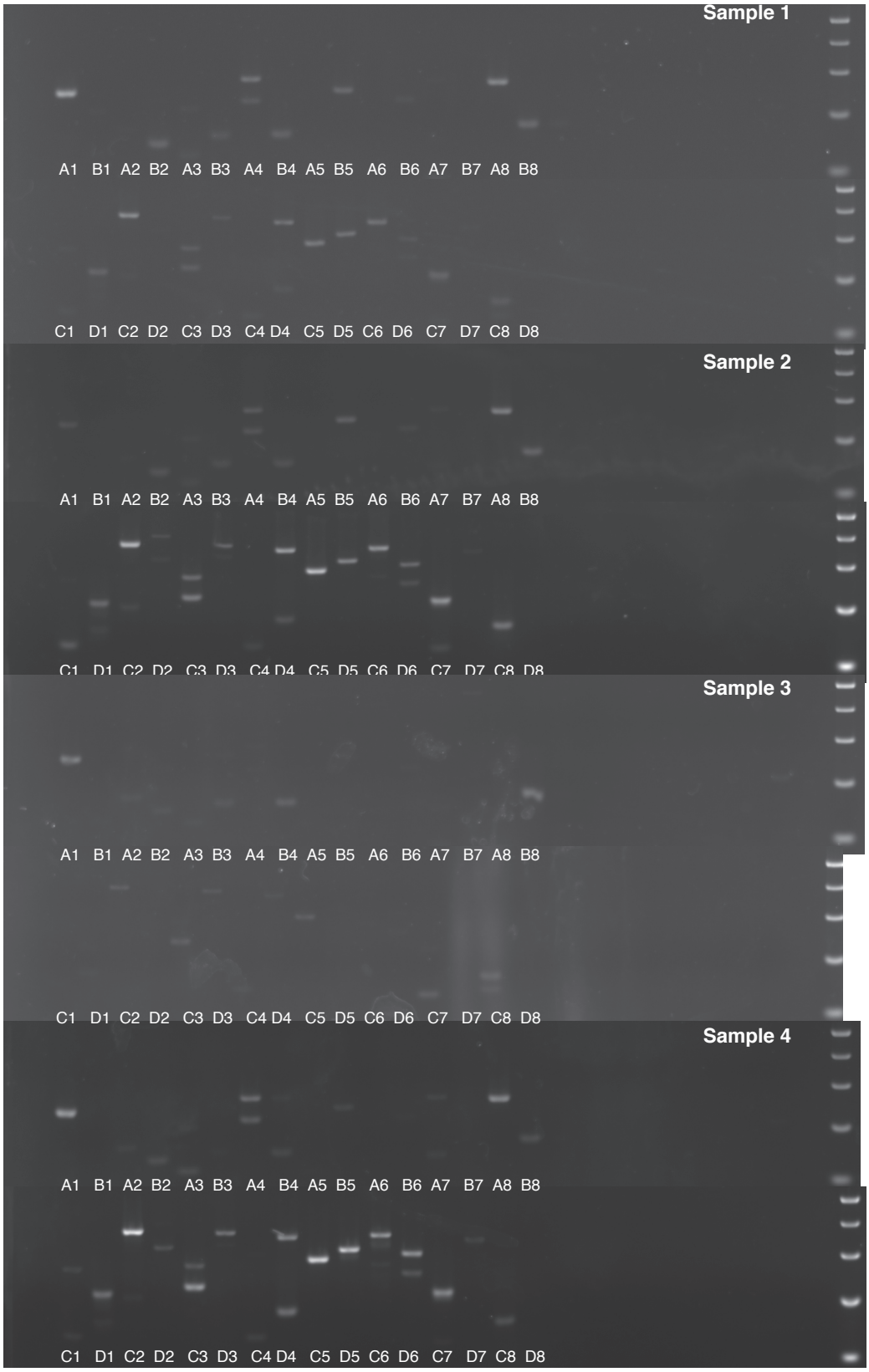


Figure 5.16: 2% agarose, 1.6 μl EtBr. SB buffer (250 V total). 45 min. 7 μl RNA sample; top two panes are + control on genomic DNA

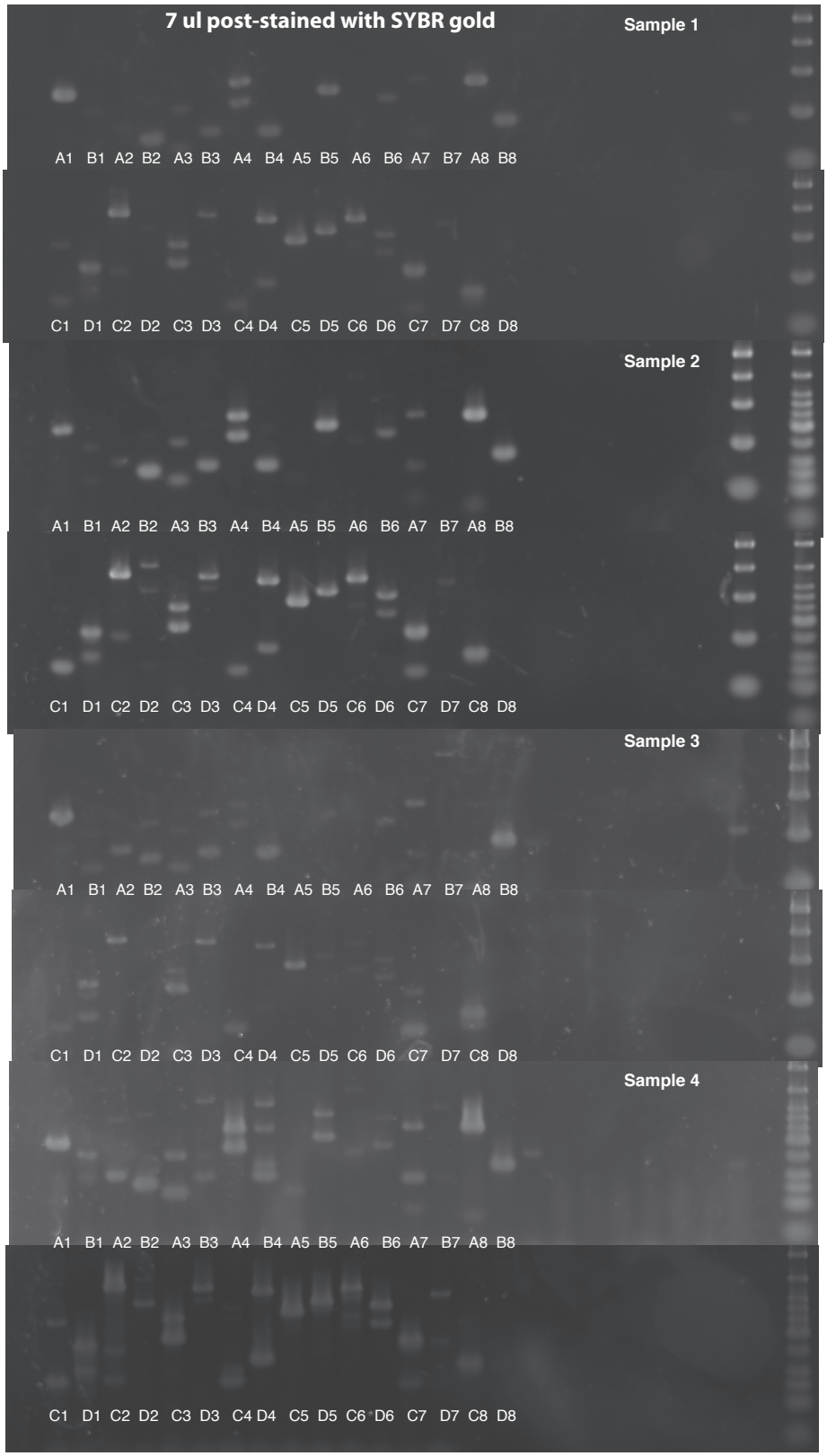


Figure 5.17: 2% agarose, 1.6 μl EtBr. SB buffer (250 V total). 45 min. 7 μl RNA sample; top two panes are + control on genomic DNA

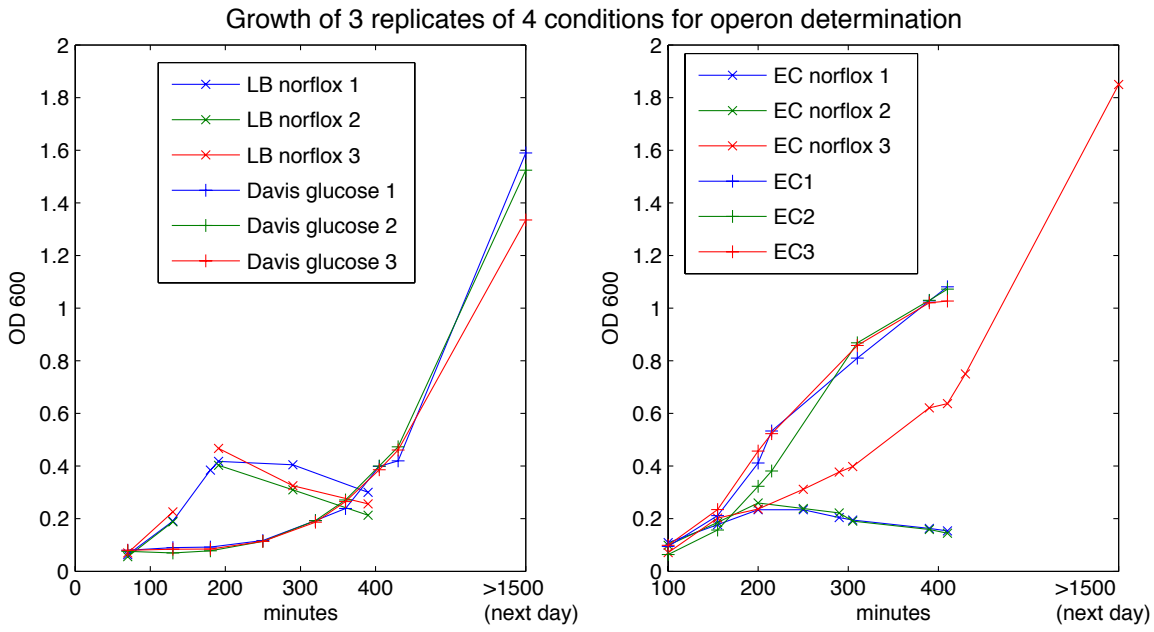


Figure 5.18: Growth curve for the 12 samples to be used for operon determination

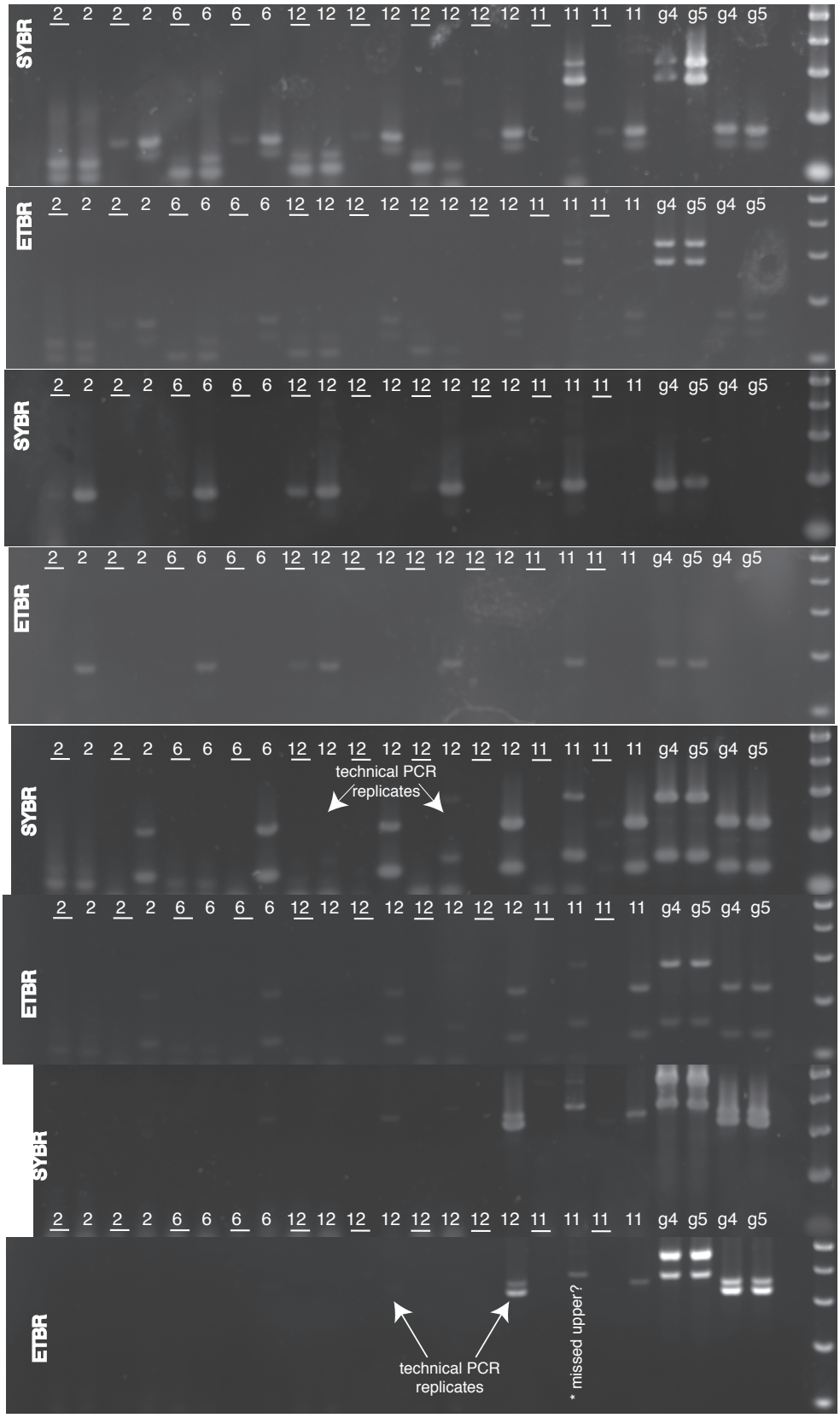


Figure 5.19: 2% agarose, 1.6 μl EtBr. SB buffer (250 V total). 45 min. 7 μl RNA sample; top two panes are + control on genomic DNA

Chapter 6

Gene and operon boundary determination by barcoded paired-end-tags and highly parallel sequencing

We want to be able to determine the transcriptional units for many species in a single sequencing run on a highly parallel sequencer (e.g. 454 or solexa).

Paired-end-tags provide an efficient mechanism to determine the 5' and 3' ends of a gene (Figure 6.18). By analyzing the sequenced tags, we should be able to get a rough idea of the transcriptional boundaries (the precise ends will probably be a little fuzzy because making full length cDNA is not always possible; Figure 6.1). Hopefully, the sequence will also provide a new means to quantify gene expression that provides a better present/absent metric than microarrays. For the first time we'll be able to determine in if all of the hypothetical genes are really transcribed. The method should also allow transcript determination/quantification in mixed cultures of undefined composition setting the stage for metatranscriptomic (Figure 6.1 Mixed Culture).

By placing an error-correcting barcode on each sample, we can take advantage of the growing capacity of highly parallel sequencers to sample multiple conditions at the same time (Figure 6.2). A barcode is just a piece of known DNA what we ligate onto our sample to identify it. For example, you could ligate AGA to the front of one some human DNA sequence and TAT to the front of some chimp DNA sequence, and then when you get the sequencing read back with AGA you know you have your human sample. The problem of course is that sequencing errors could lead to the misidentification of your sample. Because of this, I plan on using error-correcting barcodes. Error-correcting codes have been used for many years to keep things like telephones, CDs, and DVDs function properly in the prescence of noise and other problems (e.g. most CDs will work even with small stratches because there is enough redundancy built in that it can *fill in* the missing bits.

The simplest error correcting code would be if we replicated our code: AGA → AAAGGGAAA; TAT → TTTAAATTT. With this coding schema as long as you get less than one error every three base-pairs, you can recover your original sequence (e.g. AATGGGAAA = AGA). This simple method is a horribly inefficient way to do things that is good for teaching, but never used in practice. In practice, I have some software for generating a set of error-correcting DNA barcodes with a set level of misidentified barcodes (e.g. you can say you want 1 in 1,000,000,000 barcodes to

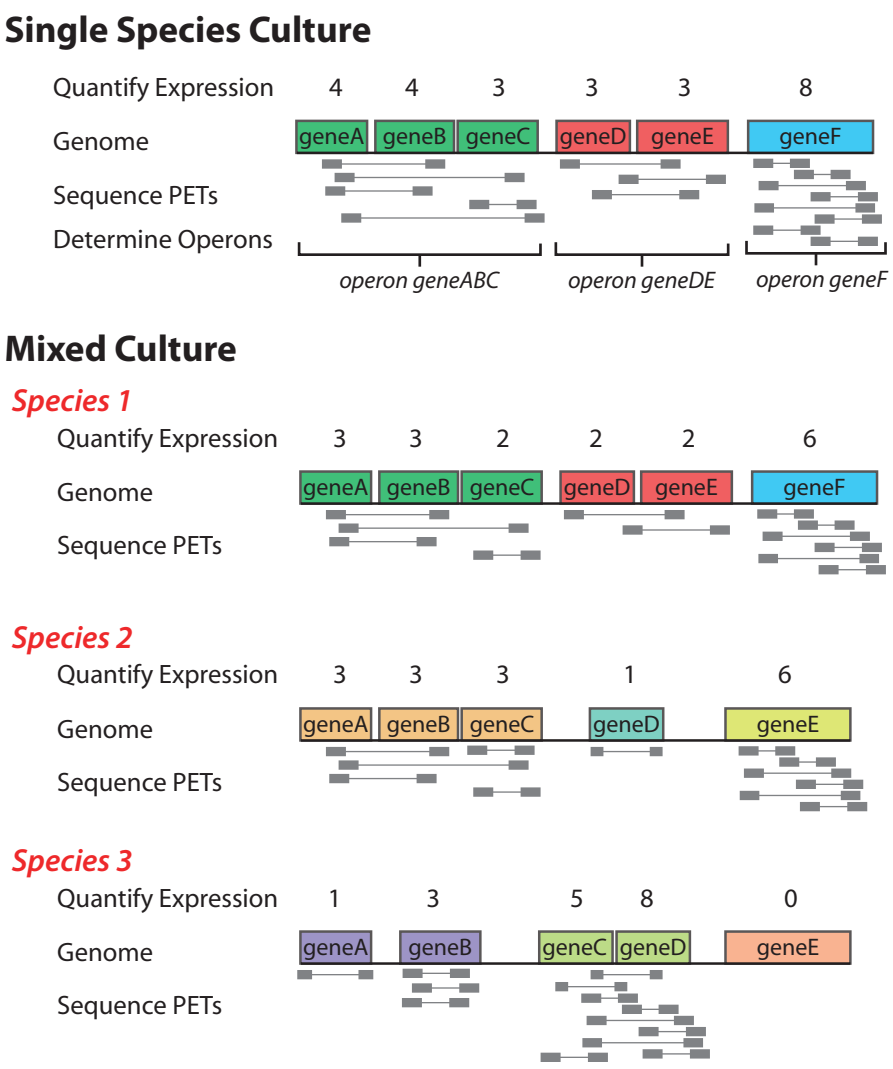


Figure 6.1: Paired-end-tags are mapped back onto the appropriate genome providing estimates of transcript boundaries (operons) and gene quantity.

fail and this places a constraint on either the length of the barcode or the number of barcodes you can make). No error correcting code is flawless, to be flawless, you'd need an infinitely long code. Since DNA has for bases = 4^N barcodes of length N, it isn't too difficult to generate code codes without using too much of your sequencing read (6-10bp).

This chapter was started:

Wed Aug 23 14:13:07 EDT 2006

6.1 Planning and Goals

1. make ds cDNA, pUC clone, TOPO clone, and sequence 10-20 (should be mostly rRNA but NOT 6S)

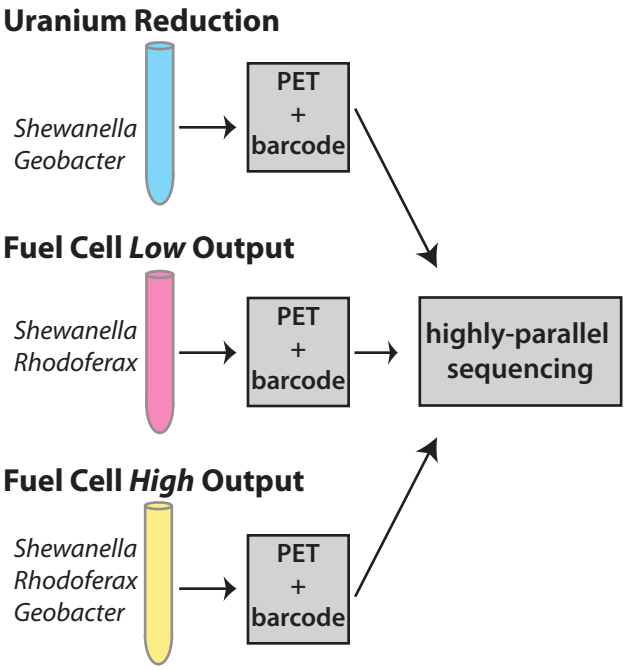


Figure 6.2: Depending on how quantitative this method is, it may be possible to use it for meta-transcriptomics.

2. make ds Cdna, added linkers, pUC clone, and sequence 10-20 (should be mostly rRNA but NOT 6S)
3. make ds cDNA, added linkers, size-select pUC clone, and sequence 10-20 (should be mostly rRNA but NOT 6S)
4. use Ambion kit to remove 16 and 22 S rRNA, make cDNA, add linkers, pUC clone, sequence 10-20 (should NOT be mostly rRNA), if rRNA is pretty rare, sequence more (see if they'll colony pick for me)
5. take rRNA mRNA, make cDNA, A-tail? or add linkers?, circularize, amplify, cut with MmeI, PAGE purify, clone tags and sequence a few
6. take rRNA mRNA, make cDNA, A-tail? or add linkers?, circularize, amplify, cut with MmeI, PAGE purify, clone tags, add linkers and sequence a few, if works sequence 96

size fractionate? http://wheat.pw.usda.gov/ lazo/methods/uo/pro1.html http://www.genome.ou.edu/protocol_bo pack beads into a spin-x column?

or use a cDNA fractionation colume from invitrogen ($29 apiece!)

6.1.1 Progress Reports on above enumeration

1. done; doesn't work very well to do TA-style cloning on cDNA. Cloning was very inefficient. All plasmids submitted for sequencing failed.
2. done, worked well. all sequences were 23S rRNA

3. done, size-selection improved the insert size. all sequences were 23S rRNA or 16S rRNA

4. transformation efficiency was very poor. Only 1 in 16 of picked white colonies had an insert. need to try again with more starting RNA

   Update Got it working better with higher transformation. Need to sequence more, but did get my first sequence that was NOT rRNA (1 out of 4; other three sequences were rRNA)

5. skipped number 5 because might was well do number 6; did get the proper 70 mer tag, just didn't want to sequence it until I put on the adaptors in step 6.

6. having troubles getting enough material to see it on a gel. I'm going to deviate a little from the Shendure at all list.

6.2 Cloning double-stranded cDNA starting from total RNA

I'm going to try and make double-stranded cDNA using slight modifications of standard approaches. In particular, I'm going to try and TA-clone the cDNA by A-tailing it with Taq and T-tailing a vector with Taq. I'm also going to try using a TOPO kit with the A-tailed cDNA. I want to use A-tailing, because that is the method employed by Shendure *et.al.* in their polony sequencing technique and will allow an easy transition from classic cDNA library protocols to a new polony protocol.

6.2.1 First steps

want >= 10-50 μg of total RNA (the binding capacity of the RNAeasy Plus column is 100 μg).

For now use 5 μg which is the max amount for the SuperScript II RT protocol.

crude protocol:

1. RNAprotect

2. lyse with lysozyme and proteinase k

3. RNAeasy

4. acid-phenol (remove more genomic)

5. LiCl (remove short RNA + clean up residual phenol)

6. remove 16S and 22S (using Ambion kit) [skip for now]

7. run 500 ng RNA on 1% gel

8. first strand 5 μg total RNA (or 500 ng mRNA) using Superscript II; use 200 units per μg of (mRNA)

9. take sample, RNA digest, ethanol precipitate, resuspend 10 μl , quantify, gel 9 μl (run all samples on same gel? 1st and 2nd strand synthesis results)

10. second strand using Sambrook and NEB enzymes

11. take sample, RNA digest, ethanol precipitate, resuspend 10 μl , quantify, gel 9 μl
12. Qiagen PCR purify (to remove short stuff and enzymes)
13. A-tag with taq and dATP (as in Shendure sequencing protocol)
14. TOPO clone
15. sequence

Detailed protocol and results:

Growing cells, RNAprotect

Wed Aug 30 10:33:56 EDT 2006

Grow 20 ml of LB with 1/100 dilution. Grow 20 ml of Davis with glucose with 1/25 dilution. Take samples in log phase. Add two volumes of RNAprotect. Vortex 5 sec, incubate at RT for 5 min, centrifuge at max rpm for 12 minutes.

Growing 6 samples 2 conditions: LB log phase and Davis minimal 0.5% glucose log-phase.

Started samples at 10:20 AM.

Italics indicates OD where samples were taken. For each condition I used 2 ml of culture and 4 ml of RNAprotect.

OD of cultures for 6 cDNA growth conditions						
min	**Davis A**	**Davis B**	**Davis C**	**LB A**	**LB B**	**LB C**
40	0.099	0.101	0.101	0.035	0.035	0.037
90	0.109	0.105	0.111	0.161	0.17	0.177
120	0.113	0.107	0.113	0.302	0.323	0.317
145	0.116	0.114	0.118	*0.433*	*0.467*	*0.457*
170	0.121	0.119	0.122	0.676	0.74	0.721
255	0.189	0.196	0.206	1.29	1.346	1.325
320	0.307	0.322	0.336	1.373	1.606	1.577
375	0.494	0.514	0.523	1.661	1.61	1.665
395	*0.498*	*0.523*	*0.552*	1.44	1.683	1.624
470	0.832	0.887	0.9	1.553	1.73	1.666

Raw data in excel format.

DNA free RNA prep

I really want zero DNA (or at least completely degraded DNA). I'm going to use the RNAeasy kit, which uses a DNA binding column to get rid of genomic DNA. Then I'm going to use the TURBO DNA-free kit. Then I'm going to acid-phenol purify the solution (acid phenol moves DNA to the organic phase). Last I'm going to do LiCl precipitation, which does not precipitate the DNA (or RNA less than 200 bp).

Samples were randomized before RNAeasy preps as:

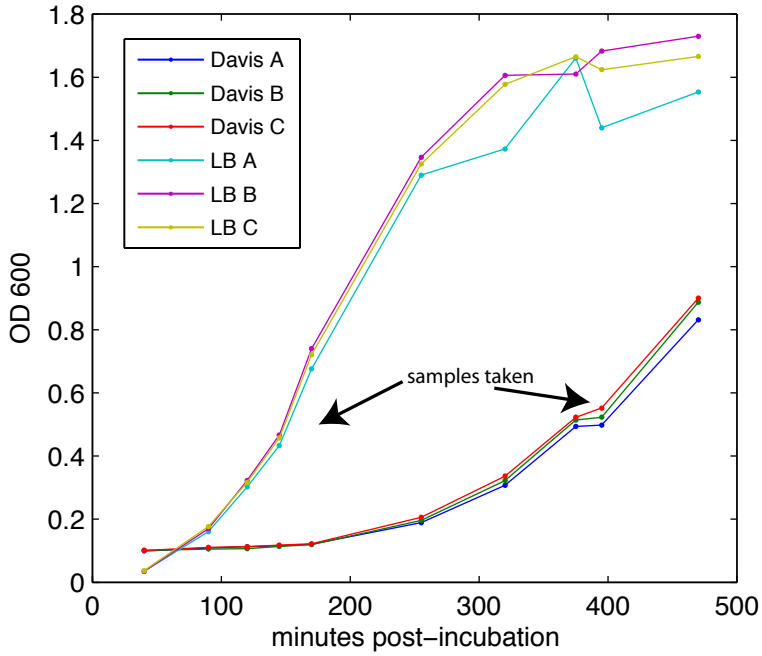


Figure 6.3: Growth curve for the 6 samples to be used for making ds cDNA

Condition	Original ID	Randomized ID
Davis and 0.5% glucose sample A	1	6
Davis and 0.5% glucose sample B	2	1
Davis and 0.5% glucose sample C	3	3
LB sample A	4	2
LB sample B	5	5
LB sample C	6	4

schema (much of this comes from the Ambion TOTALLY RNA kit manual):

1. Lyse cells in 100 μl of TE with 1 mg/ml lysozyme. Incubate 2 min, vortex every minute. Add 10 μl Proteinase K. Incubate 3 more minutes, vortex every minute.

2. add 350 μl RLT (with β-ME added) and follow the RNAeasy kit; elute with 50 μl 2 times (100 μl total)

Sample	DNA (ng/ul)	260/280	260/230	total yield
sample 1	560.8	1.97	2.16	56.1 μg
sample 2	538.7	2.22	1.48	53.9 μg

3. follow DNA-free TURBO kit instructions for high-conc DNA. Briefly: add Buffer, add 1 μl DNAse, incubate 30 min, add additional 1 μl DNAse, incubate 30 more minutes. Deactivate and keep supernatant.

4. add 200 μl TE. add 1/10 volume of sodium acetate. mix well. add 1 volume of Acid Phenol. Vortex 1 minute. Centrifuge 3 minutes at 12000 x g.

5. transfer the upper, aqueous phase to a new eppy tube

6. add 1 volume of isopropanol place at -20 C for 30 minutes
7. resuspend in 50 μl of TE
8. add 25 μl (1/2 volume) of 7.5 M LiCl; place at -20°C for 30 minutes. centrifuge at max rpm for 15 minutes
9. wash pellet in 1 ml 70% ethanol, resuspend in RNAse free H_2O .
10. resuspend in 30 μl of RNAse free H_2O .

Sample	DNA (ng/ul)	260/280	260/230	total yield	super-DNA removal loss
sample 1	638.1	1.13	1.18	19.1 μg	66%
sample 2	347.8	2.08	2.48	10.4 μg	81%

I saved 600 ng of each to run on a gel (see Figure 6.4).

Brief Conclusions: RNA prep pretty good. Will know better after I run a gel. Using 2 ml of 0.5 OD gave a pretty good yield. I'd probably even bump it up to 2.5 next time, since I'm still only half way to maxing out the column (which holds 100 μg). Loss from the DNA-removal stuff was pretty high (more than half). It isn't clear how much is due to initial DNA contamination and how much is due to all the manipulation. I'd guess most is from all the processing I did. Also, I used LiCl to remove the small RNAs. I know the Qiagen kit claims they remove them for you, but I think I probably loss quite a bit of small stuff at the LiCl stage too. Last the RNA was quite a bit dirtier after the processing. I think this is probably left over LiCl (hopefully not left-over phenol).

Brief Update *Sat Sep 2 21:49:37 EDT 2006*: I should have saved some of the initial post-genomic-removal RNA. The RNA certainly looks degraded (Figure 6.4), but I don't know where it became degraded. The 23S and 16S rRNA aren't present and the RNA smear looks like there's quite a lot of degradation. However, the cDNA lanes look longer than the rRNA that they are derived from, so maybe the RNA degraded after the cDNA was made from it. Next round I don't believe I'll use the acid-phenol step. I may move the LiCl step before the DNAse step. Next time I hope to start with more rRNA so hopefully I can use 5 μg instead of 3.5 μg

First strand synthesis of cDNA

Use Superscript II and the corresponding protocol:

Do in PCR tubes:

1. add 1 μl of random hexamers (100 ng)
2. add 1 μl of dNTP (10 mM each)
3. add 3.5 μg RNA [1]
4. add H_2O to 12 μl
5. heat to 65°C for 5 minutes, chill on ice, brief centrifuge

[1] I originally wanted to use 5 μg , but sample 2 wasn't concentrated enough for me to do that and still be under the max volume of the standard Superscript II protocol (20 μl).

6. add 4 μl First-strand buffer, 2 μl DTT
7. incubate at 25°C for 2 minutes to bind random primers
8. add 1 μl of SuperScript II mix by flicking tube a few times
9. incubate at 42°C for 50 minutes
10. heat-inactivate at 70°C for 15 min

I saved 600 ng of each to run on a gel (will add RNA cocktail before I run gel). I guessed that this would be 3.4 μl but this assumes perfect efficiency.

kept on ice while adding second strand components

Brief Conclusions: No problems will know better after gel.

Brief Update *Wed Sep 6 16:54:26 EDT 2006*: The 1st strand cDNA bands are too faint to really look at. Next time I should either not run this on the gel or run more (Figure 6.4).

Second strand synthesis of cDNA

Do in same PCR tube as first strand; no need to clean up the first strand. Keep on ice while preparing.

1. add 66.15 μl of H_2O
2. add 10 μl of NEBuffer 2
3. add 3 μl dNTP mix (10 mM each)
4. add 5 μl *E. coli*DNA polymerase I (40 Units)
5. add 0.25 μl RNAse H (1 Unit)
6. incubate 2 hours at 16 C
7. add 5 μl ligase buffer
8. add 1 μl DNA ligase
9. incubate 15 minutes at 16 C
10. heat inactivate both enzymes 20 min at 75 C
11. add 5 μl RNAse cocktail and incubate 30 min at 37 C
12. cleaned up with Qiagen PCR clean up; eluted into 35 μl EB buffer [2]

[2] 35 μl was chosen because it allows 1 μl to be used to spec the DNA and the remaining amount is the maximum allowable volume for the end-repair kit

13. spec'd DNA with 1 μl

Sample	DNA (ng/ul)	260/280	260/230	total yield	% RNA RT to DNA
sample 1	19.7	1.88	3.40	689.5 ng	19.7%
sample 2	25.8	2.08	2.98	903 ng	25.8%

14. end repair with epicenter kit using 34 μl cDNA (all of it; just keep the same tube); incubated at RT 45 min

15. heat deactivated enzymes 70 C for 10 min

16. cleaned up with Qiagen PCR cleanup; eluted into 30 μl EB buffer

Will run 600 ng on gel.

The DNA was spec'd just before blunt cloning:

Sample	DNA (ng/ul)	260/280	260/230	Yield
ds cDNA Sample 1	32.8	2.70	2.61	984 ng
ds cDNA Sample 2	29.3	2.94	3.36	879 ng

Brief Update *Sat Sep 2 21:49:37 EDT 2006*:

Brief Conclusions: The % RNA RT to DNA yield measures the amount of ds cDNA at the end relative to the 3.5 μg starting concentration of RNA. I'm not sure what this number is supposed to be. It doesn't feel too bad. However, the real % yield is actually slightly more than half this number since the 2nd strand synthesis if it were 100% efficient would double the amount of DNA from the 1st strand synthesis. RNA of course is only single-stranded.

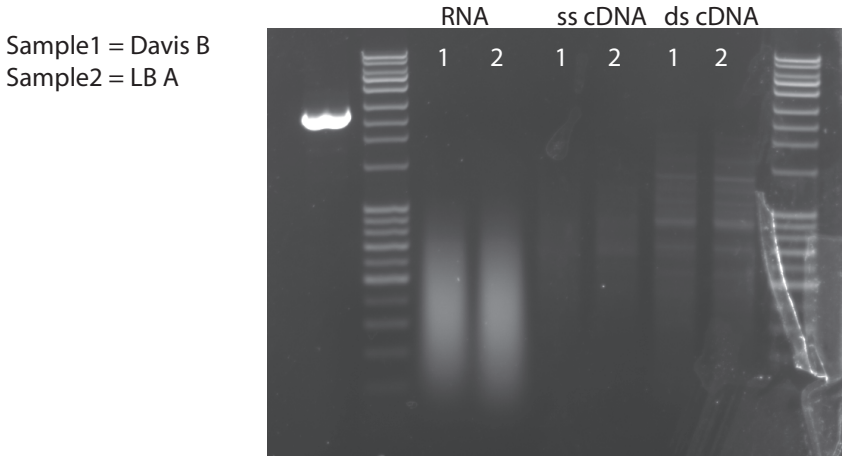


Figure 6.4: RNA, single stranded DNA (ss cDNA), and double stranded cDNA (ds cDNA). The RNA looks pretty degraded, but the cDNA doesn't. Perhaps the RNA degraded after production of the cDNA. 1 μl of RNAse cocktail was added to the ss cDNA appx 30 min before running on a gel. Gel is 1%, 0.5 cm agarose run for 60 min at 80 V (8V/cm) in TAE with 0.5 μl of EtBr.

Preparing vector DNA

pUC19 from NEB. Cloned into DH5α, so I can have an infinite supply.

Cut 5 μg of plasmid (the DNA purchased from NEB) with SmaI at RT for 45 min. Heat deactivated 20 min at 65 C. Cleaned up with Qiagen PCR cleanup. Eluted into 30 μl . Yield:

Sample	DNA (ng/ul)	260/280	260/230	Yield
pUC19	137.1	1.86	2.35	4.11 μg

T-tailing the blunt vector Wed Sep 2, 2006

18.25 μl cut vector (2.5 μg) was combined with 5 μl PCR buffer 25.2 μl H_2O , and 1 μl dTTP (100mM). This reaction is supposed to very inefficiently add T's to the end of the sequence. I don't think it worked very well. An alternative strategy is to use terminal tranferase and add ddTTP which is more efficient. The T-tailing reaction was placed at 72°C for 90 min in a thermocycler.

The reaction was cleaned up with a Qiagen PCR clean up kit and eluted into 30 μl of EB buffer. The yields:

Sample	DNA (ng/ul)	260/280	260/230	Yield
pUC19 T-tailed	67.2	1.82	2.53	2.02 μg

Cloning cDNA

Sat Sep 2, 2006

Blunt cloning the cDNA 0.5 μl vector (68 ng), 0.5 μl antarctic phosphatase, 1.7 μl phosphatase buffer, 15 μl H_2O was incubated for 15 min at 37 C. The enzyme was heat-inactived for 5 min at 65 C. 2 μl T4 Ligase buffer and 1.5 μl of end-repaired cDNA (30-45 ng) were added followed by 1 μl of T4 ligase. The mixture was ligated for 2 hrs and heat deactivated for 10 min at 65C.

A-tailing the cDNA Sat Sep 2, 2006 Taq polymerase efficiently adds a single A nucleotide to the 3' end of a double-stranded DNA piece.

Spec of A-tailed cDNA:

Sample	DNA (ng/ul)	260/280	260/230	Yield
A-tailed ds cDNA Sample 1	19.2	5.08	5.60	576 ng
A-tailed ds cDNA Sample 2	19.9	3.42	5.61	597 ng

The TA-cloning reaction was done just like the blunt by first removing the phosphates and then adding the cDNA and ligase. However the TA-ligation was only for 15 minutes (if I had to do it over again, I'd do it for 2 hrs too).

For both tranformations 2 μl of the ligation mixture was placed on ice for 15 min. Followed by a 30 sec heat shock at 37 C. Cells were placed on ice for 2 min and 250 μl SOC was added before growth at 37 C for 45 min. 150 μl was plated (after x-gal and IPTG was added to the plates).

Brief Conclusions: Only blunt Sample 1 and Sample 2 produced colonies. Sample 2 looked like satellite colonies and in the end none of sample 2 colonies grew in ampicillin.

TOPO cloning the A-tailed cDNA Sun Sep 3, 2006

I used some of the leftover reagents from the pTrcHis TOPO kit [Invitrogen] to clone the A-tailed cDNA according to the Invitrogen protocol with TOP10 cells.

Brief Conclusions: Many more colonies from the TOPO kit than from Blunt cloning. Still don't know if they have proper inserts.

Picking clones

The 2 TA-transformations had no colonies.

16 clones were chosen from the plates that had colonies (4 of each sample for both the TOPO kit and the Blunt cloning). The 4 clones from sample Blunt 2 didn't grow in LB and were presumed to be satellite colonies. I minipreped the 12 that grew.

Yields were:

Miniprep yields from cDNA first attempt clones						
Sample ID	**ng/uL**	**A260**	**260/280**	**260/230**	**Constant**	**Yield (ug)**
B1 a	315.07	6.301	1.95	2.09	50	15.7535
B1 b	311.39	6.228	1.95	2.05	50	15.5695
B1 c	268.3	5.366	1.95	2.06	50	13.415
B1 d	306.51	6.13	1.95	2.05	50	15.3255
T1 a	269.41	5.388	1.98	2.17	50	13.4705
T1 b	269.18	5.384	1.99	2.2	50	13.459
T1 c	283.73	5.675	1.98	2.15	50	14.1865
T1 d	237.22	4.744	1.98	2.19	50	11.861
T2 a	210.88	4.218	1.97	2.18	50	10.544
T2 b	321.16	6.423	1.99	2.19	50	16.058
T2 c	262.86	5.257	2.01	2.22	50	13.143
T2 d	236.64	4.733	2	2.18	50	11.832

Raw data in excel format

I meant to digest the 12 plasmids with EcoRI and BamHI, but I think I messed that up (see Figure 6.5).

Checking inserts by PCR

Thu Sep 7 11:45:41 EDT 2006

Because I'm not sure if I stuck the proper enzymes in my digestion to check for an insert (Figure 6.5), I'm going to PCR amplify the inserts. I was expecting the inserts to be realitively small anyways, so PCR is a better screening method. However, the primers just arrived today so I didn't have an opportunity to do this before.

I'll use 0.5 μl of insert (about 125 ng) and 200 nM of primer for 30 cycles with an annealing temperature of 52°C.

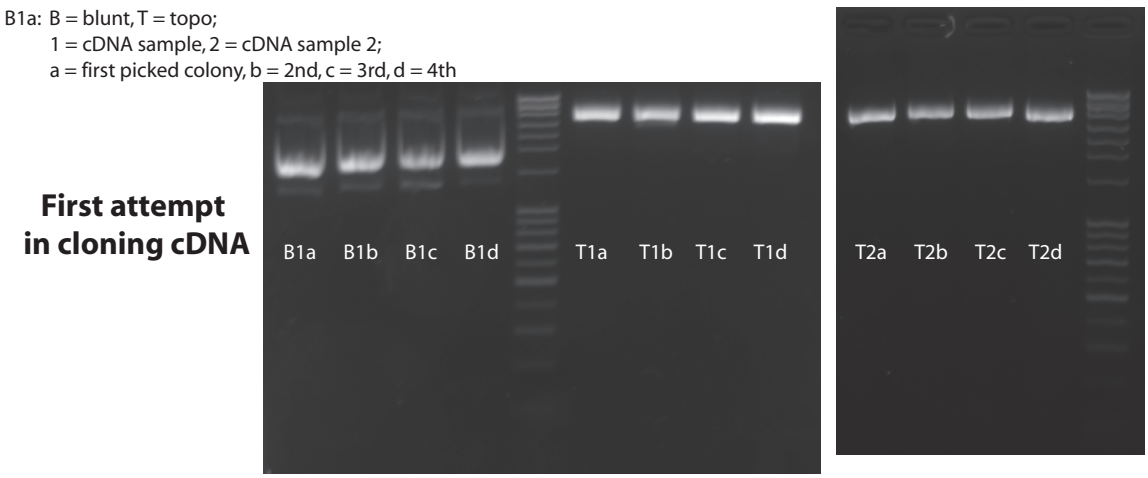


Figure 6.5: 1.5% gel pUC19 and pTrcHis TOPO vector digestion. Now that I think about it, I don't know what enzyme I used. I'm afriad I only added EcoRI. I meant to add EcoRI and BamHI.

Sequencing clones

Wed Sep 6, 2006

Four of the Blunt clones were set to Agencourt for sequencing. B1a and B1d are being sequenced in both directions.

Sequence submission data in excel format

Brief Update *Sun Sep 17 22:37:10 EDT 2006*: All of the sequences failed. Because the later projects where I used adaptors worked so much better, I don't feel any strong need to resend these out or repick,miniprep, etc them again. I do have the crap they sent me. The longest sequence was 100 bp or so and didn't match to anything related to a *E. coli* gene. The remaining 3 sequences were less than 10 bp. Not much to conclude for this part of the project, except that TA-style cloning of cDNA didn't seem to work that well. And neither did blunt. I'm also getting more and more disappointed in the failure rate of Agencourts sequencing service. I think it's time for a change.

6.3 Cloning double-stranded cDNA from total RNA, using adaptors

6.3.1 RNA to cDNA

RNA prep

1. Lyse cells in 100 μl of TE with 1 mg/ml lysozyme. Incubate 2 min, vortex every minute. Add 10 μl Proteinase K. Incubate 3 more minutes, vortex every minute.
2. add 350 μl RLT (with β-ME added) and follow the RNAeasy kit; elute with 50 μl 2 times (100 μl total)

Sample	DNA (ng/ul)	260/280	260/230	total yield
sample 3	523.0	2.00	2.20	52.3 μg
sample 4	482.0	2.04	2.35	48.2 μg

I saved 1.5 μl (appx 750 ng) of each sample for a gel (see Figure 6.6). Unfortunately, I mixed up the damn tubes. Should still give a general idea though. 97.5 μl were left for the LiCl step.

3. add 50 μl (1/2 volume) of 7.5 M LiCl; place at -20°C for 30 minutes. centrifuge at max rpm for 15 minutes
4. wash pellet in 1 ml 70% ethanol, resuspend in RNAse free H_2O .
5. resuspend in 50 μl of TE [Ambion]
6. follow DNA-free TURBO kit instructions for high-conc DNA. Briefly: add Buffer, add 1 μl DNAse, incubate 30 min, add additional 1 μl DNAse, incubate 30 more minutes. Deactivate and keep supernatant.
7. transfer the upper, aqueous phase to a new eppy tube

Sample	DNA (ng/ul)	260/280	260/230	total yield	super-DNA removal loss
sample 3	619.8	1.85	2.03	31.0 μg	40%
sample 4	612.5	1.86	2.07	30.6 μg	37%

I saved 600 ng of each to run on a gel (see Figure 6.6).

First strand synthesis of cDNA

Fri Sep 8 15:30 EDT 2006

Use Superscript II and the corresponding protocol:

Do in PCR tubes:

1. add 1 μl of random hexamers (100 ng)
2. add 1 μl of dNTP (10 mM each)
3. add 5 μg RNA [3]
4. add H_2O to 12 μl
5. heat to 65°C for 5 minutes, chill on ice, brief centrifuge
6. add 4 μl First-strand buffer, 2 μl DTT
7. incubate at 25°C for 2 minutes to bind random primers
8. add 1 μl of SuperScript II, mix by flicking tube a few times
9. incubate at 42°C for 50 minutes
10. heat-inactivate at 70°C for 15 min

This time I didn't save any first strand cDNA for a gel.

[3] 8.07 μl of sample 3 and 8.16 μl of sample 4

Second strand synthesis of cDNA

Fri Sep 8 16:30 EDT 2006

Do in same PCR tube as first strand; no need to clean up the first strand. Keep on ice while preparing.

1. add 66.15 μl of H_2O
2. add 10 μl of NEBuffer 2
3. add 3 μl dNTP mix (10 mM each)
4. add 5 μl *E. coli*DNA polymerase I (40 Units)
5. add 0.25 μl RNAse H (1 Unit)
6. incubate 2 hours at 16 C
7. add 5 μl E. coli DNA ligase buffer (NOT T4 ligase buffer)
8. add 1 μl E. coli DNA ligase (NOT T4 ligase) [4]
9. incubate 15 minutes at 16 C
10. heat inactivate both enzymes 20 min at 75 C
11. this time I did not add 5 μl of RNAse cocktail, assuming instead that the RNAse H had removed enough of it to be neglegable in the spec measurements
12. cleaned up with Qiagen PCR clean up; eluted into 35 μl EB buffer [5]
13. spec'd DNA with 1 μl

Sample	DNA (ng/ul)	260/280	260/230	total yield	% RNA RT to DNA
sample 3	134.8	2.06	2.26	4.7 μg	94%
sample 4	124.0	2.08	2.38	4.3 μg	86%

14. end repair with epicenter kit using 34 μl cDNA (all of it; just keep the same tube); incubated at RT 45 min
15. heat deactivated enzymes 70 C for 10 min
16. cleaned up with Qiagen PCR cleanup; eluted into 30 μl EB buffer
17. spec'd 1 μl

Sample	DNA (ng/ul)	260/280	260/230	total yield
sample 3	137.7	2.16	2.57	4.1 μg
sample 4	133.4	2.19	2.63	4.0 μg

Will run 600 ng on gel (see Figure 6.6). [6]

[4] if I had to do it over again, I'd add another 0.25 μl of RNAse H here

[5] 35 μl was chosen because it allows 1 μl to be used to spec the DNA and the remaining amount is the maximum allowable volume for the end-repair kit

[6] I didn't take the sample to run on the gel until after the adaptor ligation step below.

Brief Conclusions: Yield is *much* higher than last time, I don't know if that is because I got better at this or if there is still a lot of RNA that I'm measuring? I did start with almost 2x as much RNA (5 μg). Maybe last time the RNA was too degraded for a good yield?

6.3.2 Preparing cDNA and vector for cloning

Ligation of adaptors to blunt cDNA

Fri Sep 8, 2006

Ordered BamHI adaptors from IDT. One of them was ordered with a phosphorylated 5' end (the other didn't to keep them from ligating together). Ordered 110 nmole scale. The adaptor sequence is:

```
BamHI adaptor 5' GATCCGAATCCGAC
                     GCTTAGGCTG-p 5'
```

Melting temperature is around 33°C. I resuspended each to be at 500 μM, which corresponds to 1.6 μg /μl of the short piece and 2.1 μg /μl of the long piece. I combined 20 μl of each, and placed them in a thermocycler at 60 °C for 2 minutes. After the initial 2 minutes, I programmed the thermocycler to drop the temperature by 0.5°C every 30 seconds until it reached 4 °C; then I transferred the annealed oligos to ice. I should be careful not to melt the annealled oligos with my fingers since the MT is lower than human body temperature. I'll use 2 μl in each reaction (appx 4.2 μg).

1. to the 29 μl of cleaned up DNA (1 μl was used to spec), add 3.6 μl T4 DNA ligase buffer
2. add 2 μl (appx 4.2 μg) of BamHI adaptor
3. add 1 μl of T4 DNA ligase
4. mix by flicking the tube a few times
5. incubate for 12 hrs at 16°C [7]
6. heat inactivate T4 ligase at 65 C for 10 min
7. add 1 μl of T4 DNA ligase buffer[8]
8. add 1 μl of T4 polynucleotide kinase (no need to add ATP because it is in the ligase buffer)
9. incubate at 37°C for 30 minutes
10. heat inactivate for 20 minutes at 65°C

[7] Note: this didn't go exactly as planned. After 12 hours the temperature went on hold at 4°C (i.e. I didn't immediately heat inactivate the ligase, but 4°C should've slowed it down a bit). The sample remained at this temperature for approximately 5 hours before I went to the next step.

[8] this shouldn't be necessary, but I was afriad that perhaps the ATP would've been exhausted from the long ligation

11. clean up with Qiagen PCR purification kit, elute into 30 μl

Sample	DNA (ng/ul)	260/280	260/230	total yield	post-adaptor gain
sample 3	162.3	2.00	2.40	4.9 μg	18%
sample 4	156.3	2.03	2.39	4.7 μg	18%

Annealed linkers would look like:

```
BamHI adaptor 5' GATCCGAATCCGACGTCGGATTCG
                    GCTTAGGCTGCAGCCTAAGCCTAG 5'
```

GAT CCG AAT CCG ACG TCG GAT TCG GC TTA GGC TGC AGC CTA AGC CTAG

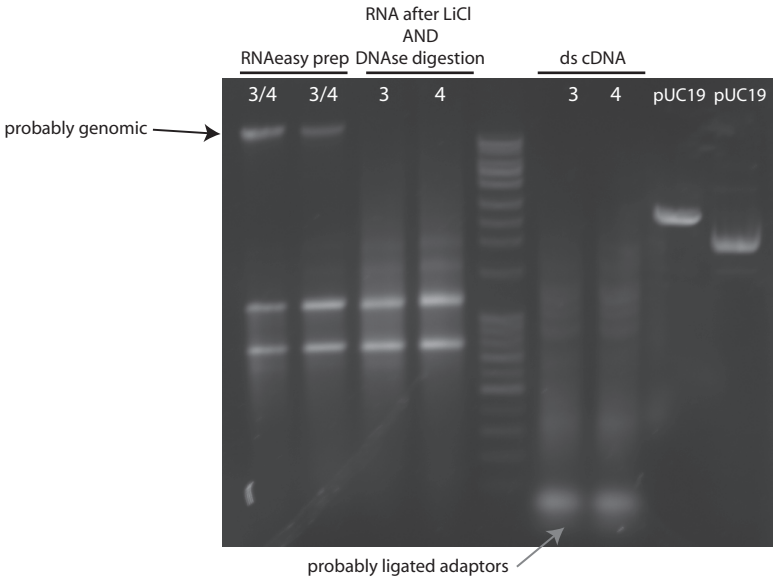


Figure 6.6: cDNA and RNA samples 3 and 4 (see section 6.2.1 on page 241 for condition and growth details). 1.0% gel, 80V(8V/cm), 0.5 cm, 70 minutes, 0.5 μl EtBr. Approximately 750 ng is in each of the RNA or cDNA lanes. The second pUC19 is the BamHI-cut vector used in the ligation.

Brief Conclusions: I'm a little worried about the gain in DNA quantity from the ligation of adaptors. The gain of 18% would indicate that the original sequence was only around 54 bp in length (a 12 bp linker is 18% of a total of 66 bp). An alternative problem would be if the annealed linkers, which are around 20-28 bp long depending on how you want to look at it, got through the column even though it is supposed to eliminate short stuff. This could happen, since I'm really loading a lare amount (4 μg) of linker through the column. The amount gained in DNA quantity would represent 80% of the short oligos being washed through. If it is the case that 20% of the primers (appx 800 ng). The gel seems to indicate that this is the case (Figure 6.6. I'm afraid all the clones are going to be this little fragment. I think I have enough to gel select (or column select, but the gel for the columns hasn't arrived yet).

Digestion of vector

Sat Sep 9 22:05:12 EDT 2006

Miniprep pUC19:

Sample	DNA (ng/ul)	260/280	260/230	total yield
sample A	104.4	1.92	2.16	3.1 μg
sample B	91.9	1.91	2.16	2.8 μg

Digested 23.5 μl of pUC19 sample A (appx 2.5 μg) with 0.5 μl BamHI (10 U) at 37 C for 45 minutes. Heat deactivated 20 min (I don't know if this helps, since BamHI can't really be heat deactivated).

Added 3.4 μl phosphatase buffer and 1 μl antarctic phosphatase to dephosphorylate for 15 minutes at 37 C followed by a 5 min heat inactivation. Reaction was cleaned up with Qiagen PCR kit:

Sample	DNA (ng/ul)	260/280	260/230	total yield
sample A post digestion	50.8	2.02	2.43	1.5 μg

6.3.3 cloning adaptored cDNA

Ligation of adaptored cDNA to cut vector

Sat Sep 9 22:05:12 EDT 2006

7.5 ng cDNA with 50 ng vector and 50 ng cDNA with 50 ng vector

Did in standard way like in the appendix. 2 μl ligation, 30 min on ice, 20 sec heat shock at 42 C, 2 min on ice, add 250 μl SOC, incubate 37 C rpm 225 for 60 min, plate.

Plated 150 μl with blue-white solution.

Brief Conclusions: Looks like I definitely have a lot of empty vector sequence. The blue/white ratio is horrible (see Figure 6.7). There are inserts, but they are swamped by no insert vectors. I think next time I should digest the vector longer, clean up the reaction in between the digest and the phosphatase, and run the phosphatase reaction longer (1 hr).

Insert checking

Mon Sep 11 11:49:26 EDT 2006

Last night I picked 8 white colonies (2 from each plate) to check for an insert. I'm going to Miniprep them, and then PCR the insert region of the plasmid DNA.

Ran 0.5 μl with 1 μl of 5 uM primer mix. MT = 52 C. Ran 30 cycles. Ran 9 μl on a gel (10 μl including the dye).

Brief Conclusions: The vectors lengths look better than I guessed. Yes, there were a lot of blue colonies. And looking at the gel for the totalRNA $\rightarrow$ cDNA (Figure 6.6), there was so much adaptor sequence, I assumed that all the inserts would be annealed adaptor sequence only. looks like 3a B and 3b A and 4b B should be sequenced. I might include 4b C and 3b B just to see what I get.

Brief Update *Mon Sep 18 00:29:45 EDT 2006*: These four were sent out for sequencing. Results will be in tomorrow (hopefully).

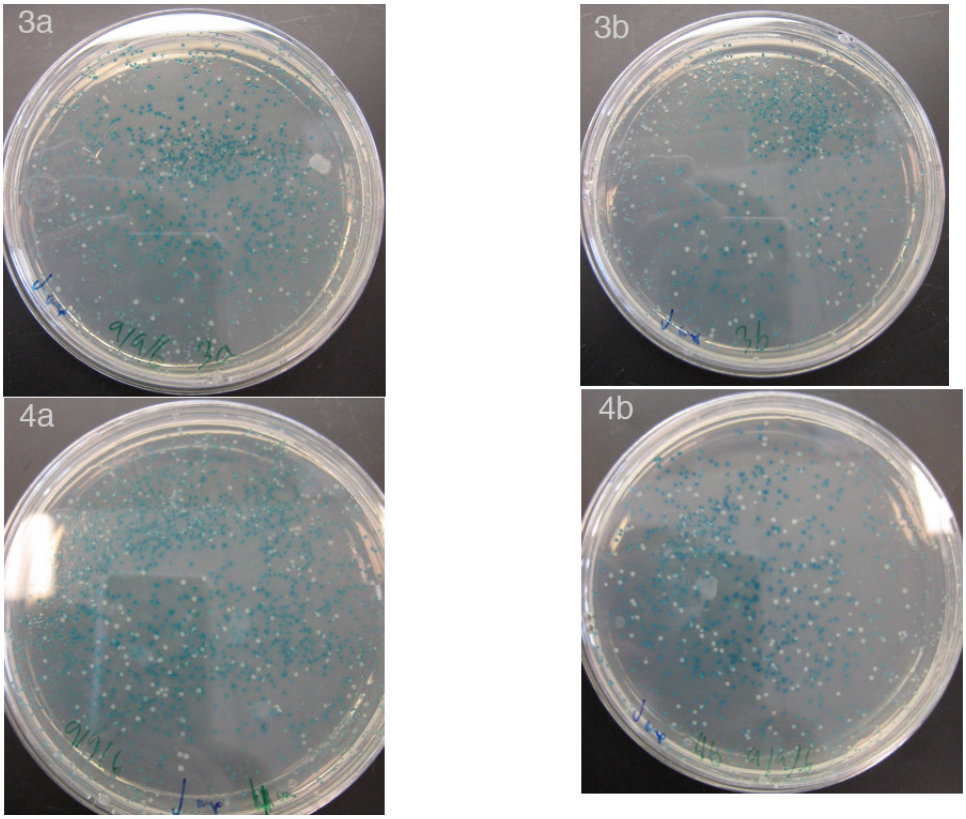


Figure 6.7: Transformation of cDNA samples 3 and 4. A = 7.5 ng of cDNA; B = 50 ng of cDNA.

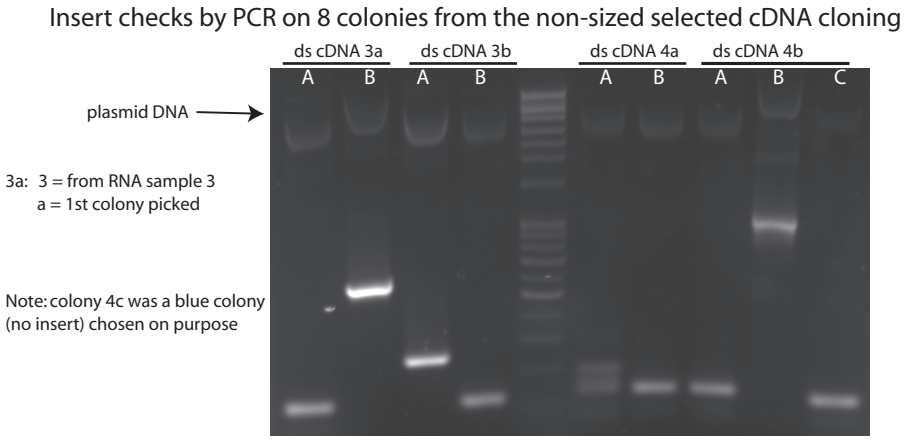


Figure 6.8: 1.5% gel 100 V (think was too hot; bands are kinda droopy) pUC19 vector amplification by PCR with M13-FOR and M13-REV. Plasmid DNA is also visible on the gel and the shift in plasmid size correlate well with the insert amplification size.

Sequencing non-size selected inserts for samples 3 and 4

(got sequences back Mon Sep 18, 2006)

I spec'd the 5 samples sent out for sequencing. 2.5 μl of each of them was sent in the sequencing reaction along with 1 μl of 20 uM forward primer.

Sample ID	ng/uL	A260	260/280	260/230	blastn (nr) result
3a B	259.12	5.182	1.94	2.18	rrlD 23S
3b A	361.66	7.233	1.93	2.22	pUC19
3b B	281.49	5.63	1.94	2.2	pUC19
4b B	320.8	6.416	1.95	2.2	pUC19
4b C	172.62	3.452	2.02	2.18	pUC19

All of the sequence data and chromatagraphs from Agencourt can be found here.

Brief Conclusions: As expected everything is rRNA. I checked if the shorter ones had inserts (e.g. the blunt ligated adaptors) or if they were just empty vectors that showed up white on the plate. 4b B is particularlly weird 3bA has 117 bp that match 16S rRNA, but the fragment is short relative to the amount of vector in the sequence so it doesn't come out as the top blast hit. 3bB matches only vector. 4bB I can't figure out. 4bC looks like there's a little adaptor seqeunce and mostly just vector.

6.3.4 size selecting cDNA for cloning

better digestion of vector?

Mon Sep 11 19:57:18 EDT 2006

Digested 23.5 μl of pUC19 sample A (appx 2.5 μg) with 0.75 μl BamHI (15 U) at 37 C for 60 minutes. This is 5 U more than the last time, and 15 minutes longer. This time I'm cleaning up the reaction before I do the phosphatase. And I'm going to use less DNA for the phosphatase step (so the phosphatase doesn't have to do as much work) in addition to running the dephosphorylation for 60 minutes instead of 15 minutes.

Sample	DNA (ng/ul)	260/280	260/230	total yield
sample A post digestion	50.8	2.02	2.43	1.5 μg

I put 200 ng of this digested vector into a 10 μl phosphatase rxn for 1 hour followed by a 5 min heat deactivation. 2.5 μl (50 ng) will be used for each ligation.

Brief Conclusions: It's not clear what helped, but the extra digestion time, clean up after digestion, extra dephosphorylation time, and use of less vector in the dephosphorylation reaction certainly lowered the number of blue colonies (compare Figure 6.7 vs the newer protocol result in Figure 6.9). I'll use this new method from now on for the BamHI into pUC19 cloning.

size-selection of cDNA via agarose gel

Mon Sep 11 19:57:18 EDT 2006

Ran a 1.0% TAE agarose gel for 50 minutes at 90V. Stained with SYBR gold Cut two ranges of DNA: med = 500-1500 bp, big = 1500-9000 bp (see Figure 6.10). The sample 3 lane contained 11 μl of sample 3, adaptored cDNA (162.3 * 11 = 1796.3 μg).

The two gel slices were placed into 600 μl tubes that had a hole poked in the bottom of them with an 18 gauge needle. The 600 μl tubes were placed inside a 2 ml tube and spun at 13,000 rpm for 1 minute to macerate the gel. The macerated gel was placed in a Spin-X column (costar) and 200 μl TE was added. The column was spun for 10 minutes. 200 μl of TE was added to the 2 ml

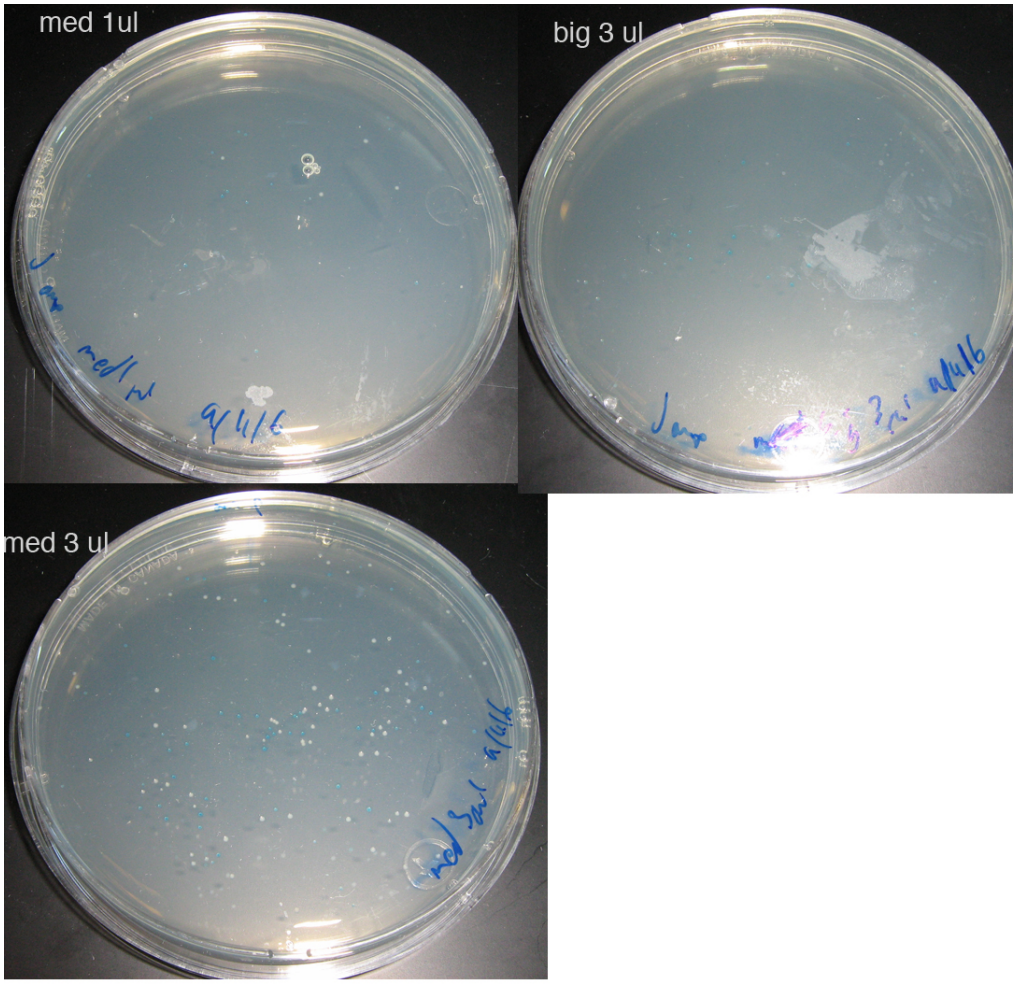


Figure 6.9: Transformation of size-selected cDNA sample 3 with a better cleaned vector. I need a better way to image these plates. Even with these poor pictures, you can see there are far fewer blue colonies (though still quite a few) than last time (see Figure 6.7). The Big 3 μl plate only had 2 white colonies, and had very few colonies total. When I replated 125 μl the next day (using most of the remaining transformation cells that were sitting on my bench), I got many more colonies.

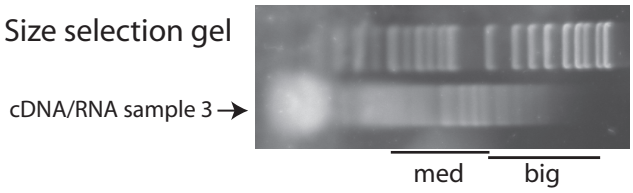


Figure 6.10: Size selection gel for sample 3. 1.0% gel 90 V TAE 0.5 cm. Stained with SYBR gold for 30 minutes, washed in H_2O for 5 minutes. Took 2 images in VersaDoc for a total of 2.5 secs under UV. Cut with a razor blade under the blue-light transilluminator (no UV). Although this image is pretty crappy, by-eye under the transilluminator it looked just fine (I've never been impressed with the versadoc images with SYBR gold).

tube (the one that had the macerated gel in it) and vortexed vigorously. This TE/gel remnant mix was added to the column which was spun another 5 minutes. 40 μl of NaAcetate was added to the spin-x'd TE/DNA mixture. 1 ml of 95% EtOH was added and mixed and the solution was placed at -86C for 15 minutes, spun for 10 min at 4 C, removed EtOH, added 1 ml 70% EtOH, spun 5 minutes at 4 C, remove supernatant. Evaporated EtOH in fume hood for 10-15 minutes. Resuspended DNA in 10 μl of TE [Ambion]. The spec readings looked really crappy (not like a mountain at all) so I wouldn't really trust these yields:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**total yield**
sample 3 *med*, post-gel cleanup	63.0	1.56	0.77	630 ng
sample 3 *big*, post-gel cleanup	72.7	1.50	0.62	727 ng

Ligation and transformation of size-selected cDNA

Mon Sep 11 20:11:24 EDT 2006

2.5 μl of dephosphorylated/BamHI cut vector was used in all reactions. 2 med reactions were run with 1 μl (3med1) and 3 μl (3med3) of the gel cleaned sample. 1 big reaction was run with 3 μl (3big3) of the gel cleaned sample (remember med = 500-1500 bp; big = 1500-9000 bp).

Ligation was 30 min at 16 C with 10 min at 65 C to heat deactivate. Transformation was as done in the previous experiment (see section 6.3.3) except that I only plated 75 μl instead of 150 μl , which gave me too many colonies last time.

The transformation worked well with plenty of white-colonies and fewer blue than before (see Figure 6.9 and section 6.3.4 on page 255).

Insert checking

Wed Sep 13, 2006

Last night I picked 16 white colonies (2 from 3med1, 10 from 3med3, and 2 from 3big3 [these were the only 3big3 white colonies]) to check for an insert. I miniprepped them, and PCR'd the insert region of the plasmid DNA.

Ran 0.5 μl with 1 μl of 5 uM primer mix. MT = 52 C. Ran 30 cycles. Ran 9 μl on a gel (10 μl including the dye).

Brief Conclusions: The vector insert lengths (see Figure 6.11) are right on target for the size-selected fragments. As expected, the med clones have insert fragments between 500-1500 bp and the big clones have inserts > 1500 bp. The number of blue clonies is way down after running a longer dephosphorylation and digestion coupled with a slightly different order of those two steps. I'm going to send several out for sequencing.

sequencing size-selected cDNA sample 3

The following five clones of sample 3 were sent out for sequencing. The two big colonies were sequenced from both ends.

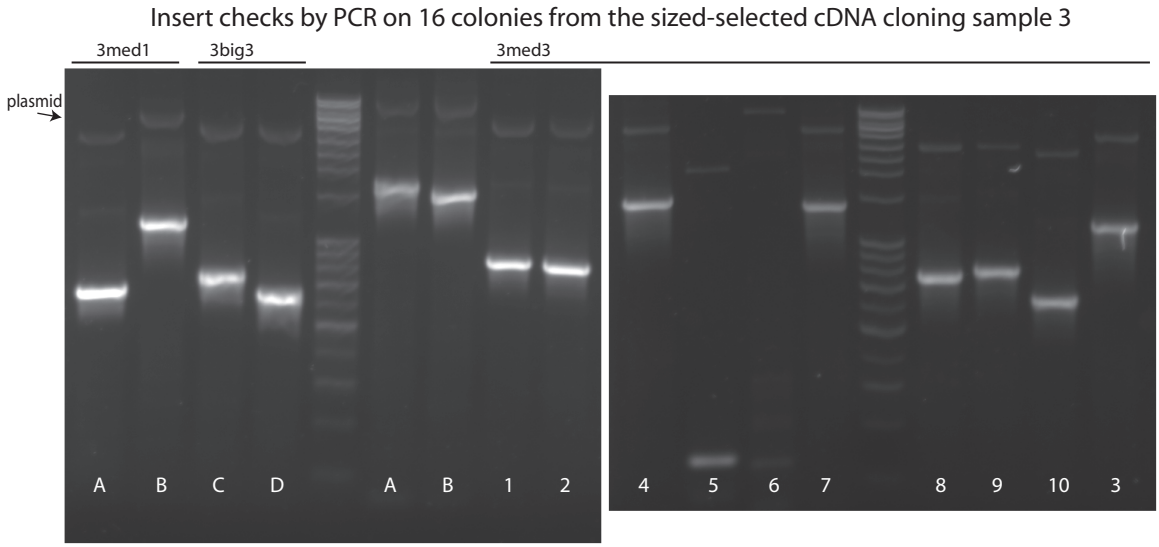


Figure 6.11: 1.5% gel 100 V (think was too hot; also think PCR melt temp was too low bands) pUC19 vector amplification by PCR with M13-FOR and M13-REV. Plasmid DNA is also visible on the gel and the shift in plasmid size correlate well with the insert amplification size. These clones were size-selected using a 1% agarose gel before cloning. *med* corresponds to clones that were from the 500 bp to 1500 bp range. *big* were clones from the 1500 bp and above (to appx 9000 bp). Except in the cases where it looks ike there was no insert, the bands all fall roughly within the correct range. Notice that the ordering is a little funky for the 3med3 as one of the tube caps broke and when I put it back on, I stuck the tube (number 3) at the end by mistake instead of in the correct place chronologically.

Sample ID	ng/uL	A260	260/280	260/230	blastn (nr) result
3med1A	102.67	2.053	1.97	2.24	rrlA 23S
3bigA	110.77	2.215	1.95	2.13	rrlG 23S
3bigB	139.18	2.784	1.92	2.21	rrlH 23S
3med3.1	135.34	2.707	1.97	2.17	rrsH 16S
3med3.4	167.73	3.355	1.92	2.25	rrlG 23S

The two big clones that were sequenced from both ends allowed me to see that only one fragment was inserted and that both of the adaptors were correctly ligated to the sequence. BigA for example had a 1579 bp insert of rrlG (see a compiled fasta file).

All of the sequence data and chromatagraphs from Agencourt can be found here.

Brief Conclusions: The good: size-selection works and helps tremendously to remove non-insert clones and short-insert clones. The bad: all rRNA. Hopefully, the MICROBExpress method for removing 16S and 23S rRNA will lessen this problem.

6.4 Cloning double-stranded cDNA from *mRNA*, using adaptors

Using samples 5 and 6 (5=5, 6=1) from section 6.2.1 on page 241.

6.4.1 RNA to cDNA

RNA prep

Tue Sep 12 10:46:43 EDT 2006

1. Lyse cells in 100 μl of TE with 1 mg/ml lysozyme. Incubate 2 min, vortex every minute. Add 10 μl Proteinase K. Incubate 3 more minutes, vortex every minute.

2. add 350 μl RLT (with β-ME added) and follow the RNAeasy kit; elute with 50 μl 2 times (100 μl total)

Sample	DNA (ng/ul)	260/280	260/230	total yield
sample 5	466.2	2.06	2.19	46.6 μg
sample 6	407.4	1.99	1.97	40.7 μg

   I saved 1.5 μl (appx 625 ng) of each sample for a gel. 97.5 μl were left for the LiCl step. Yields were a little lower than last time, hopefully the RNAprotected RNA hasn't degraded in the freezer. It's been 2 weeks (minus 1 day) since I took the samples.

3. add 50 μl (1/2 volume) of 7.5 M LiCl; place at -20°C for 30 minutes. centrifuge at max rpm for 15 minutes

4. wash pellet in 1 ml 70% ethanol incubate at RT 2 minutes, spin 5 minutes, dry pellet 7 minutes

5. resuspend in 35 μl of TE [Ambion] [9]

6. follow DNA-free TURBO kit instructions for high-conc DNA. Briefly: add Buffer, add 1 μl DNAse, incubate 30 min, add additional 1 μl DNAse, incubate 30 more minutes. Deactivate and keep supernatant.

7. transfer the upper, aqueous phase to a new eppy tube

8. spec

Sample	DNA (ng/ul)	260/280	260/230	total yield	super-DNA removal loss
sample 5	1447.5	2.08	2.28	31.0 μg	40%
sample 6	1160.3	2.08	2.24	30.6 μg	37%

   saved 0.75 μl (appx 950 ng) to run on gel

9. use MICROBExpress to remove 16S and 20S from 10 μg of total RNA (max volume 15 μl). (6.9 μl of sample 5 and 8.6 μl of sample 6) [10]

10. spec'd

Sample	DNA (ng/ul)	260/280	260/230	total yield	loss to MICROBExpress
sample 5	150.1	2.03	1.50	3.8 μg	87.7%
sample 6	166.4	2.01	1.34	4.2 μg	86.3%

[9] Last time I resuspended into 50 μl , which is better for the DNA-free kit. However, the MICROBExpress kit allows at most 10 μg in 15 μl . With 50 μl it wouldn't have been concentrated enough to get 10 μg in such a small volume. I want to make sure and maximize the starting material, because so much RNA is lost after the rRNA removal (final yield from 10 μg is expected to be 1-2.5 μg).

[10] I don't know how well this works, but it was fun to play with magnetic beads for the first time.

11. save to run on gel

Finished at: Tue Sep 12 17:56:32 EDT 2006 (7 hrs 15 min from RNA prep to here)
I saved 600 ng of each to run on a gel (see Figure).

First strand synthesis of cDNA

Tue Sep 12 10:46:31 EDT 2006
Use Superscript II and the corresponding protocol:
Do in PCR tubes:

1. add 1 μl of random hexamers (100 ng)
2. add 1 μl of dNTP (10 mM each)
3. add 500 ng of *mRNA*[11]
4. add H_2O to 12 μl
5. heat to 65°C for 5 minutes, chill on ice, brief centrifuge
6. add 4 μl First-strand buffer, 2 μl DTT
7. incubate at 25°C for 2 minutes to bind random primers
8. add 1 μl of SuperScript II, mix by flicking tube a few times
9. incubate at 42°C for 50 minutes
10. heat-inactivate at 70°C for 15 min

Again, I didn't save any first strand cDNA for a gel.

Second strand synthesis of cDNA

Tue Sep 12 10:45:43 EDT 2006
Do in same PCR tube as first strand; no need to clean up the first strand. Keep on ice while preparing.

1. add 66.15 μl of H_2O
2. add 10 μl of NEBuffer 2
3. add 3 μl dNTP mix (10 mM each)
4. add 5 μl *E. coli*DNA polymerase I (40 Units)
5. add 0.25 μl RNAse H (1 Unit)

[11] the invitrogen protocol recommends much less starting material for mRNA than for total RNA; 3.3 μl of sample 5 and 3.1 μl of sample 6

6. incubate 2 hours at 16 C
7. add 5 μl E. coli DNA ligase buffer (NOT T4 ligase buffer)
8. add 1 μl E. coli DNA ligase (NOT T4 ligase) [12]
9. incubate 15 minutes at 16 C
10. heat inactivate both enzymes 20 min at 75 C
11. this time I did not add 5 μl of RNAse cocktail, assuming instead that the RNAse H had removed enough of it to be neglegable in the spec measurements
12. cleaned up with Qiagen PCR clean up; eluted into 35 μl EB buffer [13]
13. I skipped the spec this time to save DNA
14. end repair with epicenter kit using 34 μl cDNA (all of it; just keep the same tube); incubated at RT 45 min
15. heat deactivated enzymes 70 C for 10 min
16. cleaned up with Qiagen PCR cleanup; eluted into 30 μl EB buffer
17. spec'd 1 μl

Sample	DNA (ng/ul)	260/280	260/230	total yield
sample 5	20.2	2.76	5.63	606 ng
sample 6	17.2	2.12	3.35	516 ng

Will not run 600 ng on gel as then I won't have anything left! (I'll see the cDNA when I run the gel to size select it).

Brief Conclusions: Need to aliquot more Invitrogen dNTPs.

6.4.2 Preparing cDNA and vector for cloning

Ligation of adaptors to blunt cDNA

Tue Sep 12 10:46:04 EDT 2006

I'll use 2 μl in each reaction (appx 4.2 μg).

1. to the 29 μl of cleaned up, end-repaired DNA (1 μl was used to spec), add 3.6 μl T4 DNA ligase buffer
2. add 2 μl (appx 4.2 μg) of BamHI adaptor
3. add 1 μl of T4 DNA ligase

[12] if I had to do it over again, I'd add another 0.25 μl of RNAse H here

[13] 35 μl was chosen because it allows 1 μl to be used to spec the DNA and the remaining amount is the maximum allowable volume for the end-repair kit

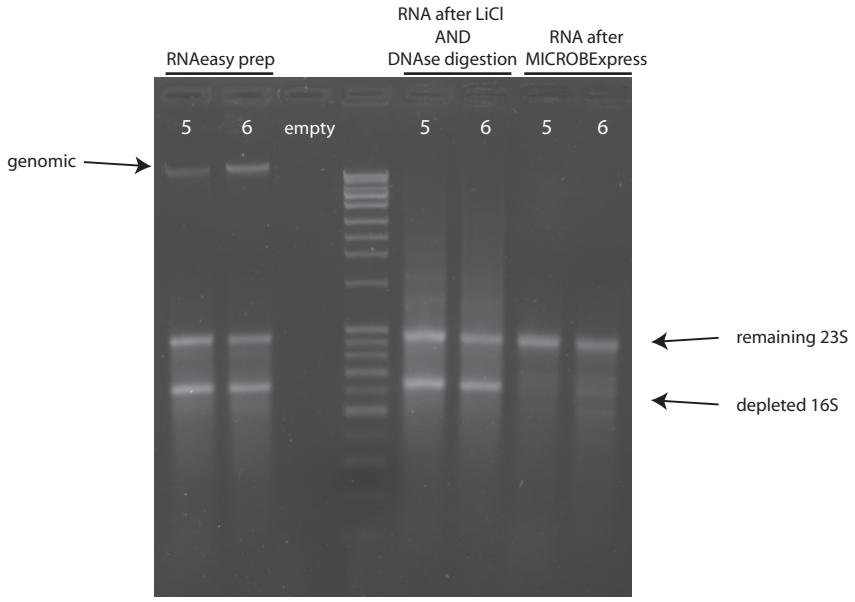


Figure 6.12: RNA samples 5 and 6 (see section 6.2.1 on page 241 for condition and growth details). 1.0% gel, 80V(8V/cm), 0.5 cm, 70 minutes, 0.5 μl EtBr. Approximately 600 ng is in each of the RNA lanes.

4. mix by flicking the tube a few times
5. incubate for 12 hrs at 16°C [14]
6. heat inactivate T4 ligase at 65 C for 10 min
7. add 1 μl of T4 DNA ligase buffer[15]
8. add 1 μl of T4 polynucleotide kinase (no need to add ATP because it is in the ligase buffer)
9. incubate at 37°C for 30 minutes
10. heat inactivate for 20 minutes at 65°C
11. clean up with Qiagen PCR purification kit, elute into 30 μl

Sample	DNA (ng/ul)	260/280	260/230	total yield	post-adaptor gain
sample 5	80.6	1.98	2.69	2.4 μg	apprx 4x
sample 6	80.3	1.85	1.59	2.4 μg	apprx 4x

Brief Conclusions: It is clear that the adaptors a certainly contributing a lot of DNA.

[14] Note: this didn't go exactly as planned. After 12 hours the temperature went on hold at 4°C (i.e. I didn't immediately heat inactivate the ligase, but 4°C should've slowed it down a bit). The sample remained at this temperature for approximately 5 hours before I went to the next step.

[15] this shouldn't be necessary, but I was afriad that perhaps the ATP would've been exhausted from the long ligation

size-selection

Performed using spin-x as in previous attempt. This time I ran all 30 μl of adaptored cDNA onto the gel. I made a 0.5 cm gel with the 6 comb. The wider lane-size allowed me to fit the entire 30 μl , but it also made it a bit tight in the spin-x tube. Because of that, I didn't do the second 200 μl TE wash of the agarose this time.

The size selection gel can be seen in Figure 6.13.

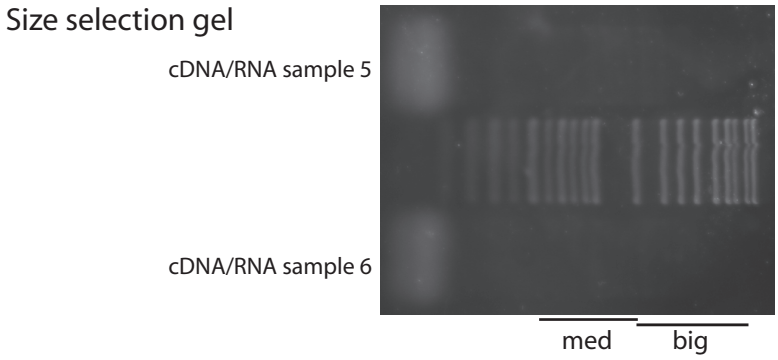


Figure 6.13: Size selection gel for samples 5 and 6. 1.0% gel 90 V TAE 0.5 cm. Stained with SYBR gold for 30 minutes, washed in H_2O for 5 minutes. Took 1 image in VersaDoc for a total of 1.5 sec under UV. Cut with a razor blade under the blue-light transilluminator (no UV). Thi is pretty crappy and by-eye under the transilluminator the cDNA was still pretty faint, so I think more 500 ng mRNA is needed in the 1st strand synthesis.

Ligation and transformation

Ligation and transformation were performed as in the previous attempt except that all 10 μl of the gel purified DNA was used. Also, the gel purification was resuspended in EB buffer instead of TE in case the EDTA present in 10 μl would've inhibited the reaction.

Insert checking

The normal PCR insert checking procedure was done with 0.5 μl of plasmid from 16 white colonies. Results are in Figure 6.14.

Brief Conclusions: None of the clones had good inserts (See Figure 6.14). And the only one that looked like it had an insert at all had a very short one (only 300-400 bp). I think the problem is not enough material. It is quite hard to clone into the BamHI cut pUC19 and not get a bunch of blue colonies. It would be much easier with a non-symmetric kinda thing like I'll use later. For now, a minor adjustment is that I might incubate with Antarctic phosphatase for even longer or add more. But more importantly I think I need to run with more starting material so that when I do the size-selection, I actually have something easily viewable to select (see Figure 6.13).

I'm also a little disappointed with the results of the MICROBExpress kit. It clearly depleted the 16S, but there still seems to be a lot of 23S hanging around. Maybe I need to run it through the magnetic bead protocol 2x (see Figure 6.12)? I should also try using a denaturing agarose gel instead of native.

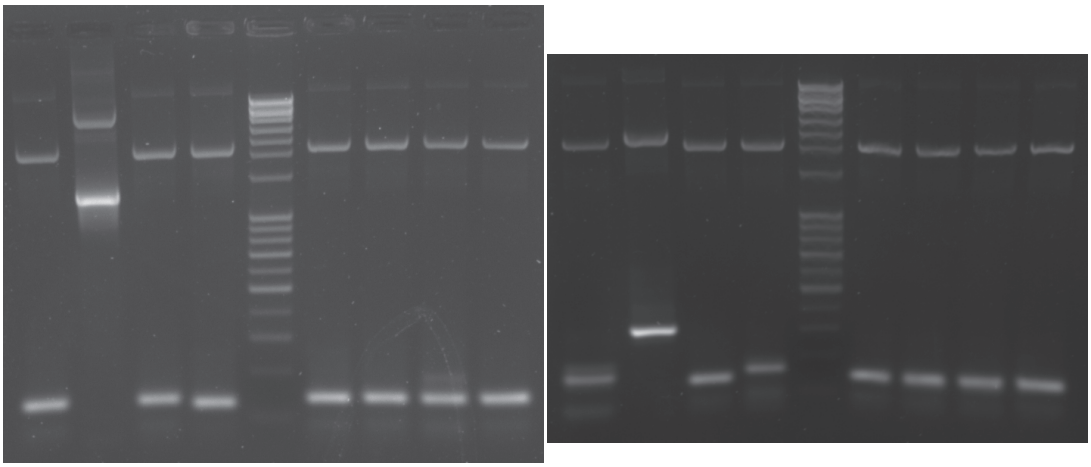


Figure 6.14: 1.5% gel. Only 2 out of 16 colonies had what looks like an insert. I need to redo the cDNA step with more RNA or use a different type a size-fractionation [like the sephacrl] or else I'm doomed with insert efficiencies like this.

6.5 Cloning double-stranded cDNA from *mRNA*, using adaptors and more RNA

Previous attempt didn't work too well with the mRNA. This time I'm bumping it up to 1.5 μg of mRNA (with is using the maximum allowable volume of 10 μl for a standard 1st strand synthesis reaction).

Most of the steps I'm doing as I did before (see section 6.4 on pages 258-263). Here are the modifications: (1) using 1.5 μg instead of 500 ng of mRNA in the first-strand reaction. (2) using 2 μl of Superscript II instead of 1 μl [16] (3) adding 0.25 μl of RNAse just prior to adding the *E. coli* DNA Ligase.

Thur Sep 21, 2006 I ran the entire sample on an agarose gel for size-selection (see Figure 6.15). It was cleaned up with a spinX column.

I cloned them in the same way as before, except that the dephosphorylation setp used 2 μl of Antarctic phosphatase instead of 1 μl and I ran the reaction for 30 minutes instead of 60 minutes. Also, I used 350 μl SOC in the transformation and plated 50 μl (250 μl and 75 μl were used in the previous attempt). I ligated 5 μl of the gel-size selected cDNA (out of 15 μl total).

I got less colonies than in the past, the white-blue ratio was bad but not horrible. I picked an initial 4 colonies (2 from 5big and 2 from 5med) to check if they were all empty like last time.

Sun Sep 24, 2006

Brief Conclusions: The inserts are actually present this time at decent lengths (compare Figure 6.16 and the previous attempt in Figure 6.14). The question now is will I have something *besides* 23S rRNA??? Based on Figure 6.12, I'm not too hopeful; Now I think the bias will be even stronger towards 23S since the kit seems to have done a nice job to get rid of the 16S!

Sat Sep 23, 2006 I plated another 100 μl the next day to have more colonies to pick.

[16] In the invitrogen manual for making cDNA they suggest 200 U (200 U = 1 μl per μg of RNA)

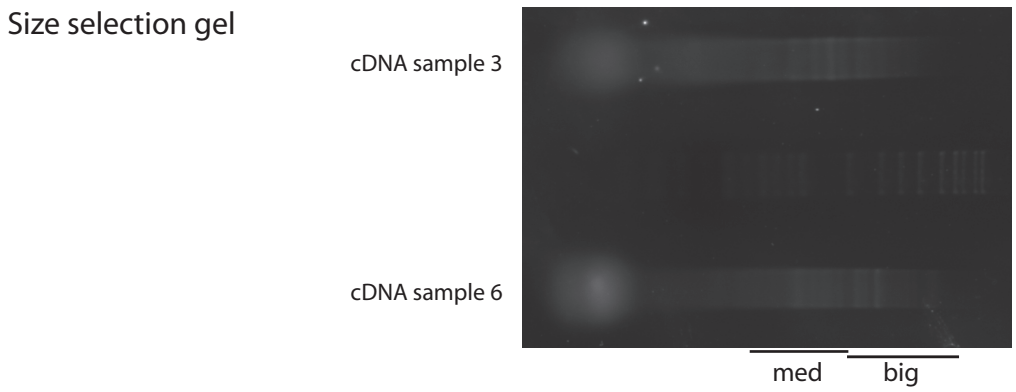


Figure 6.15: size-selection 1.0% gel for samples 3 and 5. Once again SYBR Gold visualizes crappily but it is clear as day on the transilluminator by eye. Sample 3 was size-selected for the circularization experiment below. Sample 5 was for traditional cDNA cloning.

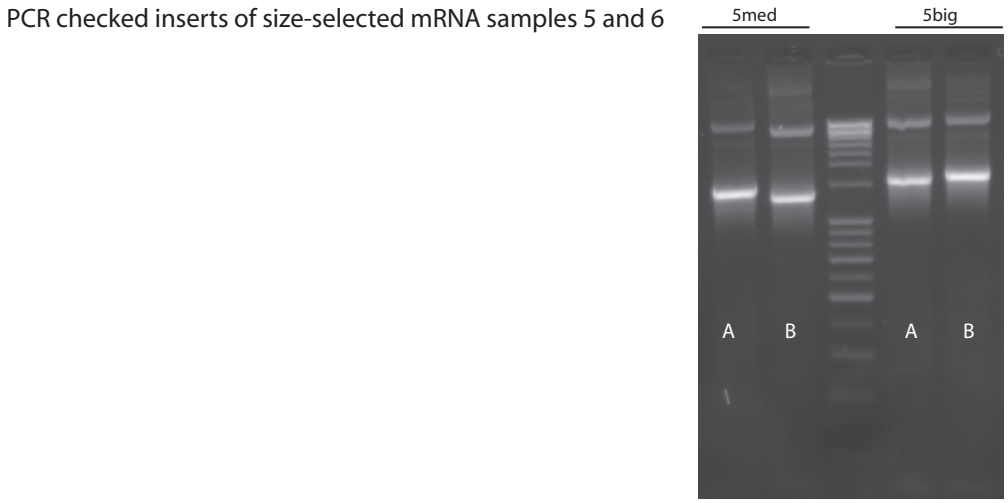


Figure 6.16: 1.5% gel EtBr. Inserts were checked by PCR using the M13 primers.

Brief Update *Mon Sep 25 15:20:44 EDT 2006*: Now that it looks like this mRNA derived cDNA is clonable, I need to sequence a bunch (10-20) and see if they are all rRNA still. I'll pick more colonies for miniprepping tonite.

I picked 16 colonies (2 from each plate in the previous plating above and 6 from each of the newer plates from Sep 23, 06). I miniprepped them and check the insert by PCR and agarose gel (see Figure 6.17).

Brief Conclusions: These additional 16 samples look good for the most part. Sample 11 is particularly interesting because the insert is more than 5kb.

Sequencing the improved insert size samples from sample 5

Thu Sep 28 15:31:49 EDT 2006

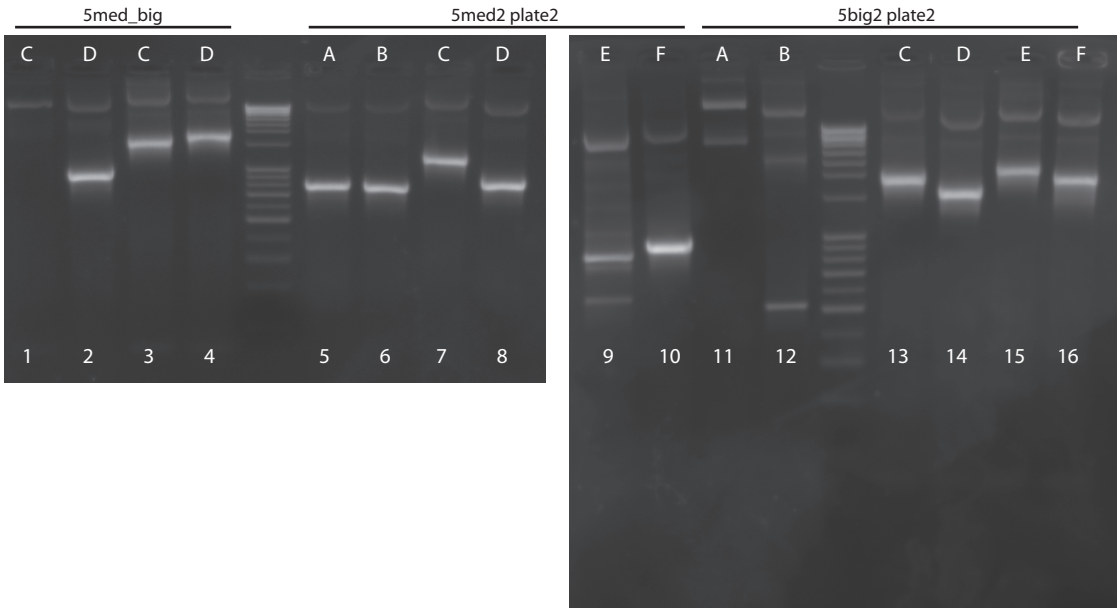


Figure 6.17: 1.5% gel sample 5 insert checks. Unfortunately, I messed up the labeling for the four samples which I picked from the Sep 23 plate. Based on the insert size I'd say that samples 3 and 4 were the *big* samples.

I sent four sequences out this morning (these were vectors checked in Figure 6.16). If they come back ok, I'll send more.

Here's the info for the sequencing:

Sample ID	ng/uL	260/280	260/230	blastn (nr) result
5big 2A	318.1	1.90	2.25	23S rrlH
5big 2B	608.0	1.82	2.15	23S rrlA
5med 2A	539.3	1.85	2.20	*tnaL* - *tnaA*!!!
5med 2B	397.8	1.87	2.15	23S rrlC

Brief Conclusions: Mon Oct 2 16:17:38 EDT 2006

As the table above shows, FINALLY I have an insert that is not a 23S or a 16S rRNA. One out of the four sequences sent was not an rRNA. Also, that one sequence is from the operon that contains tnaL, tnaA, tnaB. The sequence starts at the 41st bp of the leader sequence and proceeds to at least the 664 bp of tnaA (tnaA is 1431bp total). Rich Roberts mentioned to Simon that it may not be possible to get the 5' end of genes because the translation machinery in *E. coli* eats up (degrades) the 5' ends as it moves along. This, my first result that doesn't involve an rRNA gives at least one hint that this might not be a problem. Clearly this transcript runs across two genes, and more than that begins only 40 bp away from the start of a leader peptide not even a proper gene. Those things don't get picked up well by microarrays, but it seems that maybe this sequencing approach will catch them. Not that the sequence also did NOT read through the entire insert which is actually about 1500 bp long (see Figure 6.16). I think it would be good to sequence in the other direction, so at least I have some evidence for how long I got on this gene.

I'll probably send 10-16 more to see if I can get some stats.

Mon Oct 30 11:28:44 EST 2006

The following additional samples checked in Figure 6.17 were spec'd and sent to agencourt for sequencing. The samples had been in the fridge for about a month, so hopefully they haven't degraded.

Sample ID	ng/uL	260/280	260/230	blastn (nr) result
5med 2	239.4	1.89	2.14	rrlC 23S
5big 2B	375.7	1.87	2.17	rrlG 23S
5med 2A	127.3	1.95	2.03	kdsA + intergenic
5med 2B	228.3	1.89	2.19	rrlA + 23S
5med 2B	490.8	1.87	2.18	rrlH + 23S

Here is the raw sequence data from agencourt.

Brief Conclusions: Mon Nov 6 13:41:45 EST 2006

Now the number of rRNA to mRNA reads is 2/9. Very bad, but much better than before I used the Microbexpress kit. It is clear though that I'm either going to have to optimize that kit (e.g. annealing longer or using wather bath instead of heat block) or run the RNA through the kit 2x. However, both of the reads that I have for non-rRNA genes do yield information that would be informative to determining gene boundaries.

6.6 Circularization test with sample 3

6.6.1 Circularization adaptors and strategy

I can shorten the overhang to a 3-mer which is NOT palendromic (so that the adaptored sequences DON'T ligate to each other) by adding an extra G to the phosphorylated oligo (see section 6.3.2 on page 251 for the original adaptor).

```
BamISH adaptor 5' GATCCGAATCCGAC
                      GGCTTAGGCTG-p 5'
```

This primer should be much better than the BamHI for ligating to the circularization dsDNA oligo below, because it helps ensure that only one adapted RNA and one circularization probe ligate into a circularized piece (with the BamHI adaptor and a BamHI adaptor you could easily get concatamers of RNA or circularization oligo). All overhangs (on the adaptor and the circularization probe) will be phosphorylated to make the ligation more efficient [17].

Melting temperature is around 35°C. I resuspended the bottom at 500 μM, which corresponds to 1.69 μg /μl of the short piece and 2.1 μg /μl of the long piece[18]. I combined 20 μl of each, and placed them in a thermocycler at 60 °C for 2 minutes. After the initial 2 minutes, I programmed the thermocycler to drop the temperature by 0.5°C every 30 seconds until it reached 4 °C; then I transferred the annealed oligos to ice. I should be careful not to melt the annealled oligos with my

[17] As I've done previously with the BamHI adaptors, I'll phosphorylate the adaptors AFTER they've been ligated to the RNA. It's not really as important with the NON-palandromic adaptor, but since I already had an unphosphorylated adaptor, I'll stick with the post-phosphorlation method.

[18] the long piece is the same I used before

fingers since the MT is lower than human body temperature. I'll use 2 μl in each reaction (appx 4.2 μg).

dsDNA fragment for circularization

```
circularization adaptor   5' p-ATCGGCCAAGGCGGCCGTACG
                                  CCGGTTCCGCCGGCATGCCTA-p 5'
```

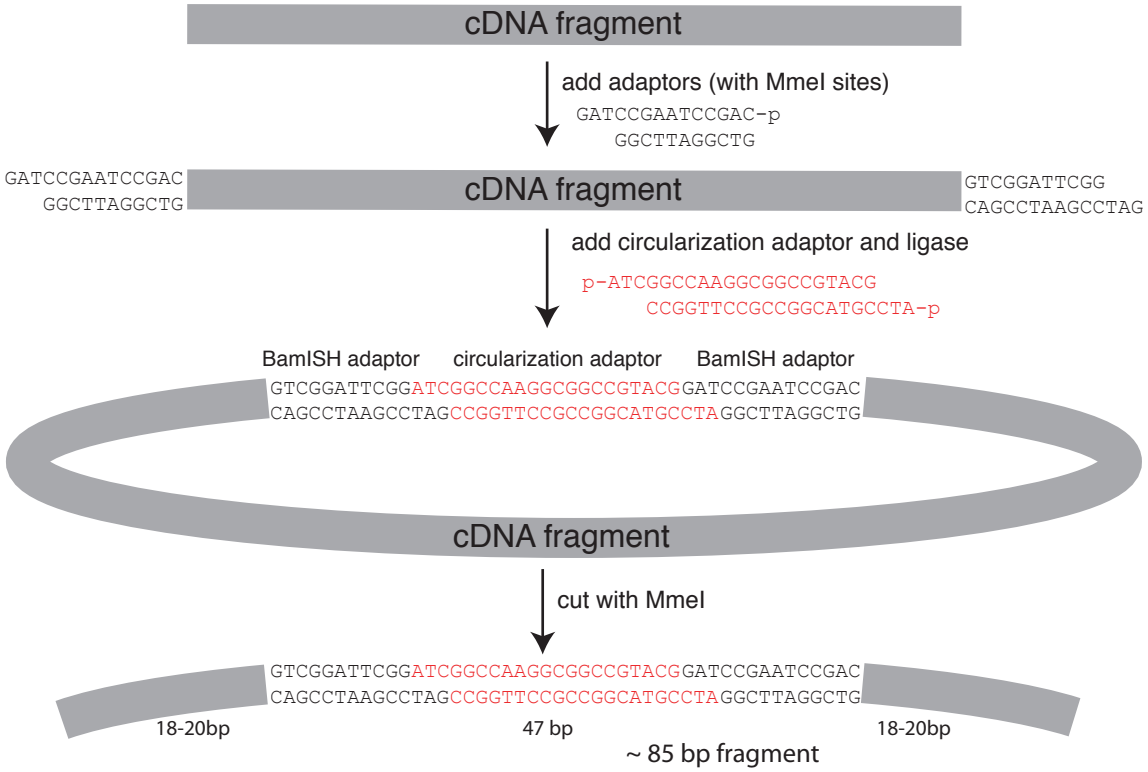


Figure 6.18: Schema of the circularization strategy.

Melting temperature is around 62.8°C.

6.6.2 Circularization first attempt

Circularization by ligation

Thur Sep 21, 2006

I used approximately 120 ng (0.5 μl) of circularization oligo. Combined with 5 μl of adaptored size-selected cDNA. I used 2 μl T4 ligase, 1 μl of T4 ligase. I ran the ligation for 2 hr at 16°C followed by heat-inactivation of the ligase.

RCA amplification and MmeI digestion

Fri Sep 22, 2006

I ran RCA according to the fidelity systems protocol. 1.5 μl template, 2.5 μl annealing buffer, 1 μl fidelity systems random hexamer primer (modified to reduce exonuclease reaction). The only modification of their protocol was running the reaction for 4 hr instead of the recommended 12 hr.

This reaction certainly amplified something because in the end (after running through a Qiagen PCR cleanup), I had decent yields of DNA:

Sample	DNA (ng/ul)	260/280	260/230	total yield
RCA 3med	43.9	1.92	3.05	1.7 μg
RCA 3big	50.3	1.99	3.16	1.5 μg

I digested 15 μl of each RCA product with MmeI. I ran this digested RCA product and 15 μl of uncut RCA product (all that remained) on an agarose gel

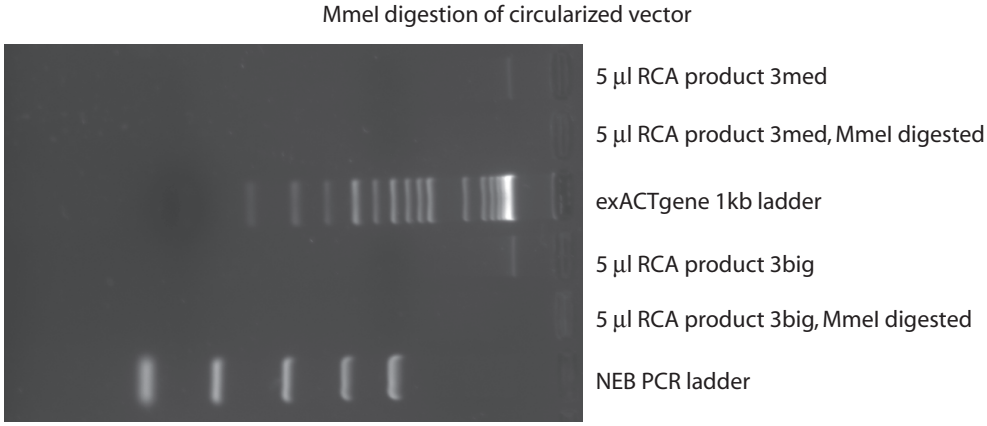


Figure 6.19: 2% agarose gel. Doesn't look like things worked. Somewhere along the way (ligation with circularization oligo or maybe RCA) the small pieces seem to have ligated into giant pieces. 15 μl is around 850 ng of DNA.

Brief Conclusions: It looks like somewhere along the way of ligation to circularization oligo, exonuclease digestion, and RCA amplification something went awry. You can see on the gel (Figure 6.19) for the uncut lanes there is one giant piece. I guess as a long shot, the circularized piece might move VERY slow because of some strange supercoiling. Actually, maybe it worked. Now that I read more about RCA, it should make giant pieces. However, in Shendure *et.al.* they cut 40 μg of the RCA material to digest with MmeI. The ran 1/4 th of this in one lane of a gel (10 μg). This is more than 10x what I used. I'm going to try the RCA again with triple the starting concentrations and run overnite. We'll see if this is any better. I'll use SYBR gold to stain it too.

RCA amplification and MmeI digestion, try 2

Tue Sep 26, 2006

I tried multiplying the previous reaction by 3 to get more DNA. I eluted into 50 μl and had the same total concentration (so an increase in 40% or so for a 3x increase in material). I ran it on sample 3med only.

Sample	DNA (ng/ul)	260/280	260/230	total yield
RCA 3med	51.9	1.86	2.60	2.6 μg

Brief Conclusions: After I digested it, I was ethanol precipitating the large volume digestion so that it would fit in a gel lane. Unfortunately, I was using a 600 μl tube that isn't the best fit in the centrifuge. In the final spin, the tube shattered. So I don't know if this worked or not. Nonetheless, the slight increase in DNA for a tripling of reagent and template is disappointing.

RCA amplification try 3

Wed Sep 27, 2006

I repeated the RCA more in line with what Jay used in his Science polony paper. They use less dNTP but quite a bit more enzyme. Per reaction I used: 5.25 μl 10x buffer, 2 μl dNTP, 5 μl template DNA, 2.5 μl hexamer, 35.25 μl H_2O . I heated at 95C for 5 minutes (I put the sample in the thermocycler when the temperature was still at RT). I cooled them down to 4C in the thermocycler and immediately transferred the samples to ice. Then I added 2.5 μl of ϕ29 polymerase and incubated at 30C for 12 hours followed by 10 min at 65C to deactivate the enzyme. I ran to of these 52.5 μl reactions using the circularized 3big sample for both reactions.

I combined the 2 RCA reactions into one tube and ethanol precipitated them (30 min -86C, 15 min spin 4C, 750 70%, 5 min spin). Resuspension was slow. I used 200 μl of EB buffer, 50C, and periodic vortexing. It still took a while (30-60 min) to go back into solution. There was clearly a *lot* of DNA. Upon initially adding the ethanol, you could see quite a lot of DNA after 30 secs or so. The yield was very high (more in line with what Jay reports in his paper):

Sample	DNA (ng/ul)	260/280	260/230	total yield
RCA 3big	650.4	1.78	2.09	130 μg

Brief Conclusions: I think the DNA from the RCA step is incredibly long. I think the Qiagen PCR prep was selectively removing most of it, so now with EtOH precipitation I see the real yield.

MmeI digestion of RCA amplification try 3

Wed Sep 27, 2006

I digested approximately 10 μg of RCA amplified sample 3 with MmeI. The 30 μl digest was: 15 μl RCA, 5.75 μl water, 3 μl Buffer4, 1.25 μl SAM, 5 μl MmeI. I didn't clean it up. After a 30 min incubation at 37 and 10 min at 65C[19], I added 5 μl dye and ran the sample on a 2% agarose gel 12V/cm for 55 minutes (see Figure 6.20).

Brief Conclusions: I think we're in business!!! The MmeI tag of approximately the correct size is visible (see Figure 6.20). The band is not so sharp because I ran the stuff on agarose, but nevertheless it is clearly visible. I think next time I should increase the volume of the reaction and increase the amount of enzyme or incubation time, because there are clearly going to be a LOT of MmeI sites. I don't know if things get messed up when the concentration is too high?

I ordered some PAGE gels so I can do a proper gel extraction without the fuzzy problem that comes with low MW DNA on an agarose gel (like in Figure 6.20). Then the only remaining step is to ligate the two big primers on the end and sequence it. I don't think I'm going to sequence these tags. I think I'll wait and sequence everything after I've added the two ends.

[19] I know the MmeI deactivation temp is 80C for 10 min, but I was worried about melting the DNA. Probably wouldn't affecting anything by going to 80C, but who knows.

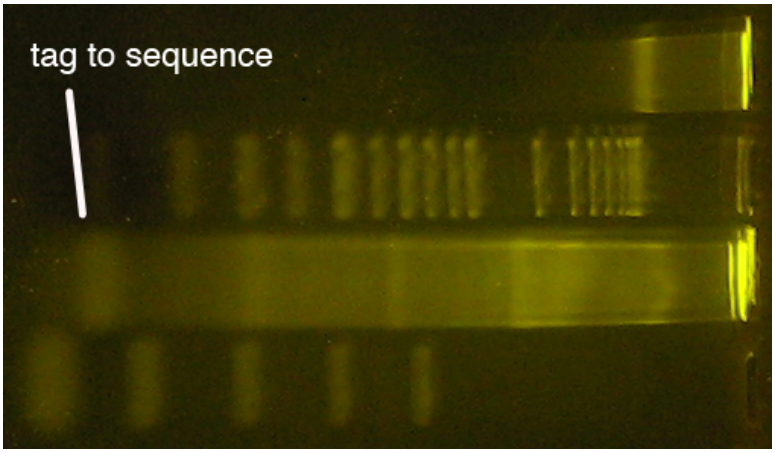


Figure 6.20: 2% agarose gel, sybr gold stained of circularized, RCA amplified, mmeI digested cDNA sample 3. I imaged it on the transilluminator with the $200 digital camera. There's a slight jiggle effect (i.e. my hand wasn't steady enough), but it is much clearer than the $30K Versadoc for some reason. The first lane 3 μl of uncut RCA DNA for sample 3. The second lane is a fisher 1KB ladder, the third is an NEB PCR ladder.

6.7 Ligating the final adaptor ends for polony sequencing

The Shendure *et.al.* polony approach uses two adaptors on the ends to sequence a few base pairs inwards on the MmeI digested piece. This is my last step then I can sequence them and I'm ready to go (though I still need to more efficiently remove rRNA).

6.7.1 Ligating for cloning using circularized/digested sample 3

I don't understand why the oligos they used in the Science paper were blunt. It seems using this strategy the adaptors can ligate in any orientation and then following PCR reaction allows you to select the properly oriented ones, but maybe I'm not understanding their method correctly? Here are their adaptors:

```
FDV2:
5' AACCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT
   TTGGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA 5'

RDV2:
5' AACTGCCCCGGGTTCCTCATTCTCT
   TTGACGGGGCCCAAGGAGTAAGAGA 5'
```

You can see there is a big length difference between them. This is because you need to gel select the ligation produces that are as follows:

```
FDV2:mmeA_circularizer_mmeB:RDV2
```

So you can use the length differences to distinguish that correct one from the ones with either two FDV2 or two RDV2 adaptors.

I'm going to add an extra few bp to the end of the adaptors to allow me to force a direction to the adaptors and to make it easy to clone them into a vector.

```
FDV2_EcoRI:
5' AATTAACCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT
       TTGGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA 5'

RDV2_BamHI:
5' GATCAACTGCCCCGGGTTCCTCATTCTCT
       TTGACGGGGCCCAAGGAGTAAGAGA 5'
```

TTGACGGGGCCCAAGGAGTAAGAGA

To try with tag: save 1 μl add phosphates? and ligate to cut puc19 save 1 μl for PCR run remaining 20 μl on gel

I think it would've been good to have phosphates on the blunt end of the oligo linkers just like with cDNA. Next time.... grow puc19

Digestion with MmeI

Tue Oct 18, 2006

As in the Shundure paper, I want to digest 40 μg of DNA (61 μl of the RCA from section 6.6.2 on page 270).

I'm running the following reaction 4x (in 4 different tubes), 15 μl RCA, 5 μl NEBuffer4, 5 μl MmeI, 2 μl SAM, 23 μl H_2O . This reaction was incubated at 37C for 30 minutes followed by heat deactivation at for 65C 10 minutes. I ethanol precipitated the digestion and resuspended in 80 μl of TE.

The maximum amount per 10 lane, 1.0mm Novex TBE gel is 25 μl . The 80 μl of cut RCA product was split into 4 and combined with 5 μl of loading dye (25 μl total) and run on a 6% TBE polyacrylamide gel. Two lanes of Low Molecular Weight DNA Ladder [NEB] were also run.

The gel was stained with SybrGold and imaged by hand using a the blue-light transilluminator (Figure 6.21).

Brief Conclusions: Perhaps it would be useful to run the digestion through a spin column, if in the future I don't do an initial gel size-selection of the 70-mer? It would get rid of the large and small pieces. The Qiagen gel extraction kit claims it retains 70bp-10kb. The PCR retains 100bp-10kb. My tags are 80-85 bp. I think the percentage TBE gel used by Shendure et al is also pretty low. Mention long pieces site this gel and previous agarose gel

To Do!!! Run the digestion again with 2x? Ethanol precipitate 1 and Qiagen gel extract the other. Run both on gel. Maybe run 3x and biotin select one?

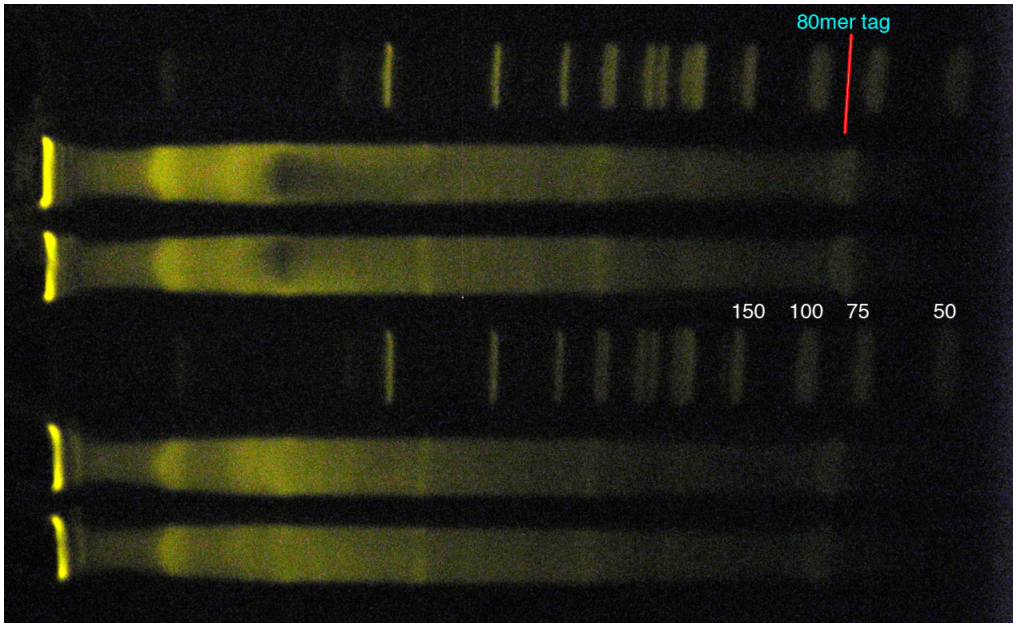


Figure 6.21: 6% TBE polyacrylamide gel showing the 80-85mer tag released from the RCA amplification by MmeI. This image has been modified using the Autocontrast feature in PhotoShop. The original can be found in the Oct, 2006 image directory.

Elution of size-selected tags from gel

Oct 18, 2006

The tags were cut from the gel with a razor blade and all four gel pieces were transferred to the same tube containing 600 μl of elution buffer [10 mM Tris-HCL (pH 7.5), 50 mM NaCl, 1 mM EDTA (pH 8.0)] (see D.2.3 on page 433 for details).

The elution was left overnight. PCP extracted and EtOH precipitated and resuspended in 20 μl of EB buffer. I was supposed to run 2 μl on a gel. I screwed up and only ran 1 μl (Figure 6.22). The gel was to quantify the yield from the gel elution. Although, I only ran 1 μl , I (and the versadoc software) could still pick up the faint band. The versadoc software estimated this band to be 8.2 ng/μl .

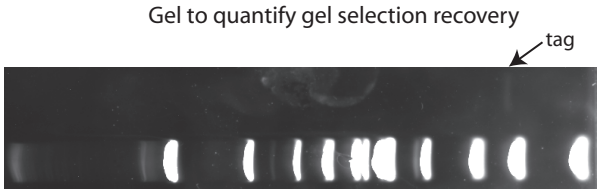


Figure 6.22: 6% TBE polyacrylamide gel for quantifying yield. 1 μl

Brief Conclusions: I gotta use EtBr with these diagnostic gels, the Sybr stuff just sucks on the versadoc. My yield was 8.2 ng/μl a little less than the 12.5 ng/μl reported in the Shendure paper.

End-repair of eluted tags

Oct 19, 2006

Used 13.75 μl of the tag (100 ng as in the Shendure paper) in a 20 μl End-Repair reaction [Epicenter]. Used 2 μl 10x buffer, 2 μl ATP, 2 μl dNTP, 0.25 μl enzyme, 0.75 μl H_2O . I heat deactivated and EtOH precipitated. Resuspended in 10 μl of EB buffer.

Ligating on the linker

Oct 19, 2006

I used 2 μl T4 buffer, 10 μl tag (all), 3 μl linker 1, 3 μl linker 2, and 2 μl ligase. I let the reaction go overnite at 16C.

Oct 20, 2006

I took 0.5 μl of the ligation and ran a PCR on it to try and enrich the correctly ligated pieces using PCR. I didn't deactivate the ligase (assuming 95C from the PCR would denature it anyways). I did deactive the remaining 19.5 μl of the ligation. The PCR reaction was 15 μl EasyA, 13 μl H_2O , 1 μl RDV2F primer (10 uM), 1 μl FDV2F (10 uM), 0.5 μl ligation rxn.

The PCR product and the 20 μl of ligation ligation were run on a 6% polyacrylamide gel (Figure 6.23).

Brief Conclusions: Obviously this wasn't the best looking result in the world (see Figure 6.23). The gel is painfully messy (again the SybrGold problem). The PCR reaction created a giant smear (Figure 6.23a fourth lane). The ligation without the PCR had a few bands, the strongest band is the tag, but it is surrounded by a smear, not by other bands. What I really wanted to see was three bands: 135 bp = RDV + tag + RDV, 157 = RDV + tag + FDV (band we want), 179 bp = FDV + tag + FDV.

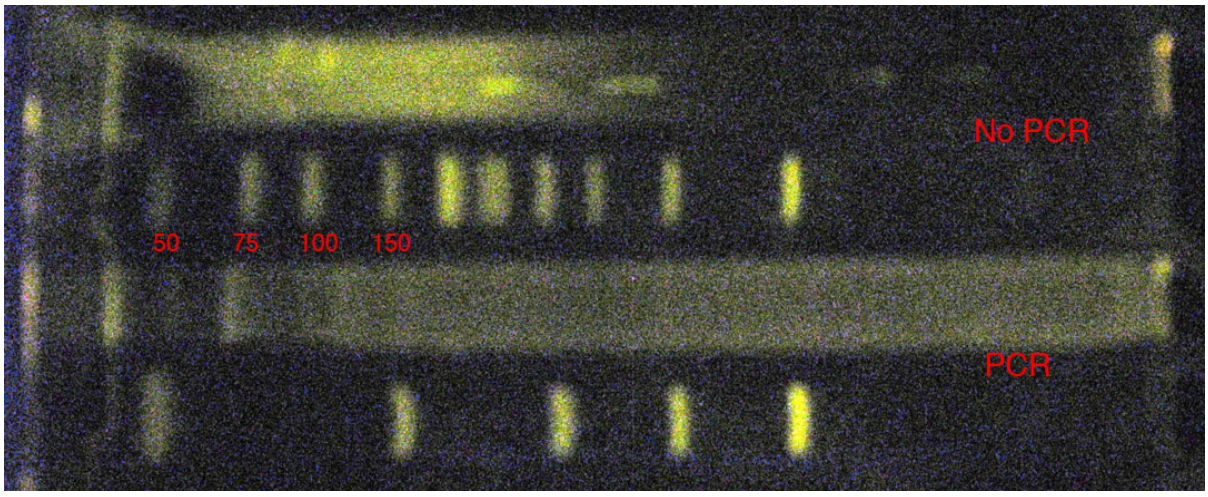
I certainly think it would be better to add phosphates to the linkers, that will vastly increase the efficiency of the ligation and will prevent having nicks in the DNA. The downside is that it will create three concatamers: RDV+RDV, FDV+FDV, and RDV+FDV. However, this is no different than the adaptors I ligated together earlier to the cDNA, and they're so short it won't be hard to pick them out (or remove them with a microcon column).

If you look close you'll see the PCR lane has 3 bands (Figure 6.23a), but the biggest band is about 20bp shorter than I'd expect.

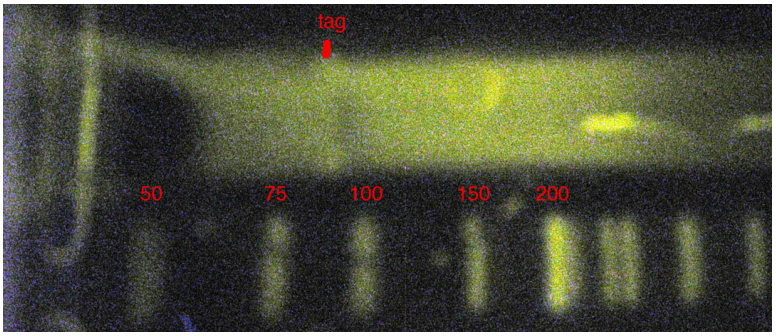
The gel was hard to photograph but here is the summary of my observations by eye: PCR product lane: giant smear with bands at 45bp (presumably the primer), another band at 65ish (presumably the ligation together of the two primers to each other, then 3 bands at 100, 120, 140 (perhaps the correct bands but they are 30 bp shorter than I thought they should be. they giant smear continues all the way to the edge of the gel

Ligation lane: the original band of 85 bp was clearly visible, but the rest was just a smear of different sized stuff from 500 bp down to 50 bp

I want to run just the dsPrimers together and see if they form distinct bands.



(a)



(b)

Figure 6.23: a) the ligation product (No PCR) and the PCR producted from amplifying 0.5 μl of the ligation product. b) zoomed in on the No PCR lane

Brief Update *Fri Nov 10 11:55:10 EST 2006*: I mentioned above that I was getting a bands at 45 and 65bp. Now I'm pretty sure those were my primers (see Figure 6.27). I don't know why yet, but the primers are migrating much slower than their dsDNA size...

Running the primers only on a gel

Oct 20, 2006

I decided maybe the smearyness (Figure 6.23) is due to not having my primers purified (which would create an assortment of lengths). ssDNA is supposed to be run on a UREA gel; I didn't have one, so I ran them (and a PAGE purified 60mer that ilaria gave me) on a 6% TBE polyacrylamide gel (Figure 6.24). The gel was stained with SYBR gold.

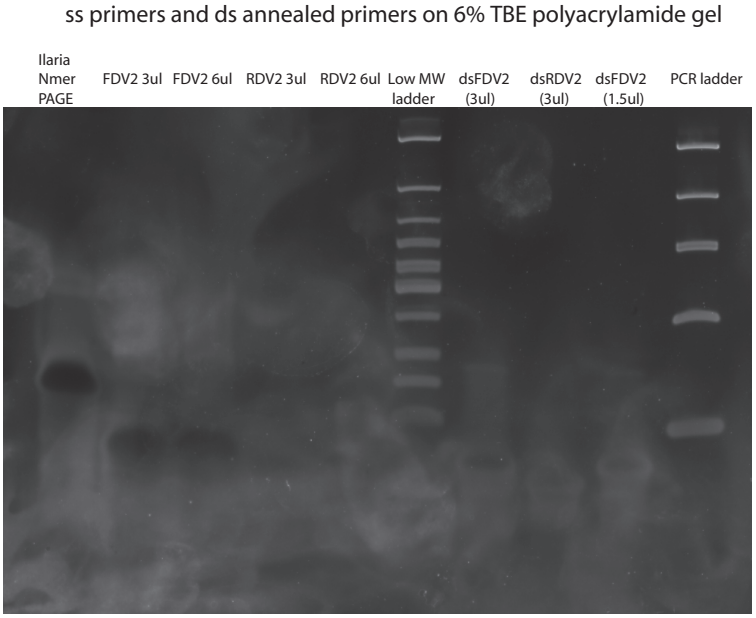


Figure 6.24: 6% TBE polyacrylamide gel for looking at ssDNA

Brief Conclusions: Once again the SYBR gold make imaging impossible on the versadoc (Figure 6.24). The bands are very smeary and look like black-hole negative signals rather than positive white signals. The black spots do seem to be in the correct location. I think the secondary structure is killing me and I need to use the UREA gel. A UREA gel might even help with detecting the ligation of the linkers????

Running the primers only on a *Urea* gel

Wed Nov 8 16:13:14 EST 2006

I'm rerunning the primers on a gel. This time I'm using a 15% TBE Urea gel [Invitrogen]. I diluted the DNA 1/2 in TBE Urea loading dye [Invitrogen]. I did *not* do this with the ladder. For the ladder, I used the normal one that I use for TBE polyacrylamide gels. This certainly didn't look like it worked 6.25. I guess I need to make a ladder with Urea dye next time.

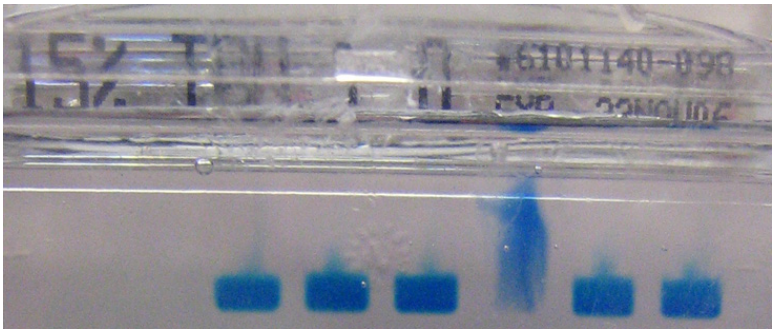


Figure 6.25: I didn't use the right kind of dye for the ladder lane. The fourth lane from the right has the wrong kind of dye and is very diffuse.

The same amounts as in the previous approach were used, except I only ran one ladder and I did not run two different concentrations of the FDV2 and RDV2. The primers were at 10 μM and I used 6 μl of each. As described in the invitrogen manual, I heated the dye/samples to 70C for 3 minutes. Then placed on ice. I also flushed the wells of the gel three times with 100μl TBE. I ran the gel at 180V for 50 minutes. I then dyed the gel for 20 minutes in 50 ml of 2 μg /ml EtBr and washed it for 10 minutes in H_2O .

The amps were a little lower than the Invitrogen protocol said they should be. But the voltage was fine.

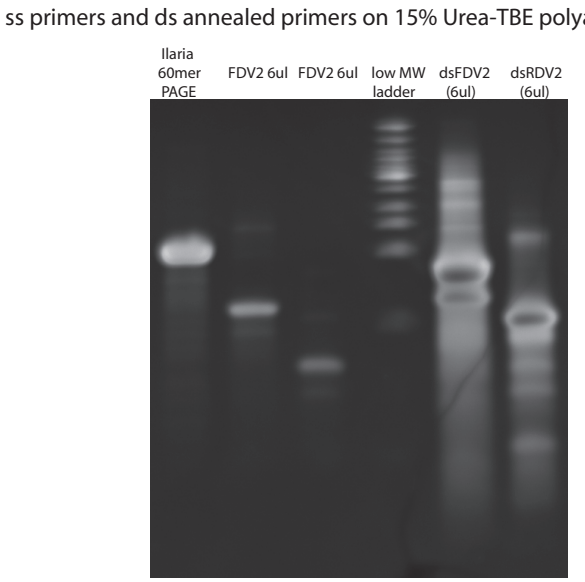


Figure 6.26: primers urea gel

Brief Conclusions: Next time use the correct kind of dye with the ladder. It looks like there is too much DNA in both the Ilaria primer lane and in the dsDNA lanes 6.26. Perhaps the "black spot" effect that I've seen several times is due to excessive DNA? Run the next gel with less DNA, especially the dsDNA. I should also run FDV2F, FDV2R, RDV2F, RDV2R, dsFDV2, and dsRDV2, just to make sure that there isn't something wrong with the reverse primers. Maybe add a little

salt to the dsDNA so that it will anneal tighter? Could probably get by with only 0.5 of ilaria's primer if I run that one again. There is some hint of impurities in the primers (shorter pieces), but it is very weak relative to the real signal

Running the primers only on a *Urea* gel, less DNA, more salt

Thu Nov 9 12:14:14 EST 2006

I'm going to try adding salt to the dsDNA to let them anneal. I ran the dsFDV2 and dsRDV2 with and without 50 mM NaCl2. I also want see if less DNA removes the black hole problem. So instead of the concentrated annealed DNA, I ran 3 μl of 10 mM forward and reverse in each lane (so the dsDNA should have the same amount of DNA as the ssDNA).

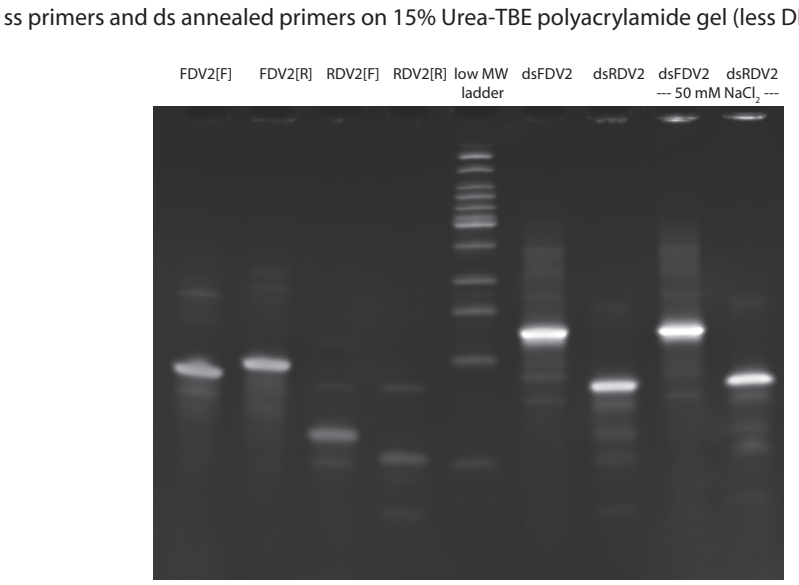


Figure 6.27: primers urea gel2

Brief Conclusions: I hate when things start to become clear, and it leads you to a point where you're not quite sure what to do. What is starting to become clear is that my annealed primers with 4bp overhands don't migrate exactly according to their length (Figure 6.27). They always move slower than their single-stranded counterparts. You might conclude that it is the double-strandedness that is slowing them down, but this isn't the case when you look at the ladder whose bands match the single-stranded primers. The 42/47bp dsDNA band is migrating at 60-65bp. The 29/25bp dsDNA band is migrating at around 45 bp. So it seems like these overhangs are slowing them down by appx 15bp? Very strange but it seems to explain two of the bands that I couldn't figure out in section 6.7.1 on page 276 (also see Figure 6.23).

It is also clear that these things are pretty smeary. Maybe purification (HPLC or PAGE) would do some good for that problem. I'm starting to favor the following idea: Do RCA, clean with EtOH precipitation, digest with mmeI, clean up in a qiagen gel cleanup column (this will remove the very small and the very large fragments [20]), Run the end repair reaction. Cleanup again with the Qiagen

[20] my piece at 80bp is a little below the removal limit for the PCR kit, but a little above the limit for the gel cleanup kit. Hopefully I don't lose too much. I can try this real easy with the RCA sample I have.

column (don't know if this too much and will remove too much DNA), spec on nanodrop, ligate on phosphorylated adaptors (each 10bp shorter than in previous round), heat deactivate (EtOH?), enrich with the dynal beads and the 30mer tag, wash off enriched DNA and concentrate in a microcon 30K or 50K (will remove short stuff), run on TBE urea gel and cut out the correct band (you could even synthesize the incorrect bands and the correct bands and test the enrichment?, would be expensive)

To Do!!! buy 30mer w/ dual biotin. buy shorter primers (17mer and 30mer) with phosphates, buy one with overhang, one without (so I can run on a gel and compare for this weird shift problem); mix dsFDV2 and dsRDV2 into same lane and run. Try nusieve? buy 50K microcon. run primers through 30K and 50K and see how it does at removing them. see: http://hcgs.unh.edu/protocol/msat/CAenrich.html for dynal instructions

Bought

Dualbiotin-spacer18-CGGATCGGCCAAGGCGGCCGTACGGATCCG appx 300 bucks!

try to anneal at 0.1M rather than the recommended 1M (natasha said it's to hard to remove the strand at that conc).

when linking use 50 (molar) fold excess? checkout http://hcgs.unh.edu/protocol/msat/CAenrich.html for linker ligation ideas too (they use 2 μl of ligase)

```
AATTAACCACTACG  FDV3   5' CCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT
    TTGGTGATGC                GGCGAAAGGAGAGATACCCGTCAGCCACTA-P 5'

GATCAACTGCC     RDV3   5' CCGGGTTCCTCATTCTCT
    TTGACGG                   CCAAGGAGTAAGAGA-P 5'
```

6.7.2 Ligating ends using a PCR product

I don't want to keep wasting my material (of which I only have a little bit left before I need to make more cDNA). Ilaria had a pair of primers for amplifying the pLtet promoter. She had two forward primers and one reverse, the first forward primer yields a 81bp product. The second yields a 120bp product. These two pieces give me a nice way to test ligating the two different ends on simultaneously in a cleaner system where I can try to optimize the ratios if they need to be optimized. And I can use them with the primers to test the removal ability of the Qiagen PCR/gel columns and the microcon columns (see section 10.3 on page 397 for more on this). Last, even though the results from the primer removal weren't that promising, having shorter primers in this next set of ligations enables me to bump up the molar ratio of adaptors without raising the quantity of DNA I use. A big problem with the previous attempt was that I had so much primer I couldn't see what was going happening on the gel.

Observing linker primers

Nov 16 and 17, 2006

I wanted to have a look and make sure I got only one band when running the primers out and not a smear. The first time (Figure 6.28a), I didn't use enough primer. The second time (Figure 6.28b), the amount was plenty for the 30mer but still to short for the 15mer with the phosphates.

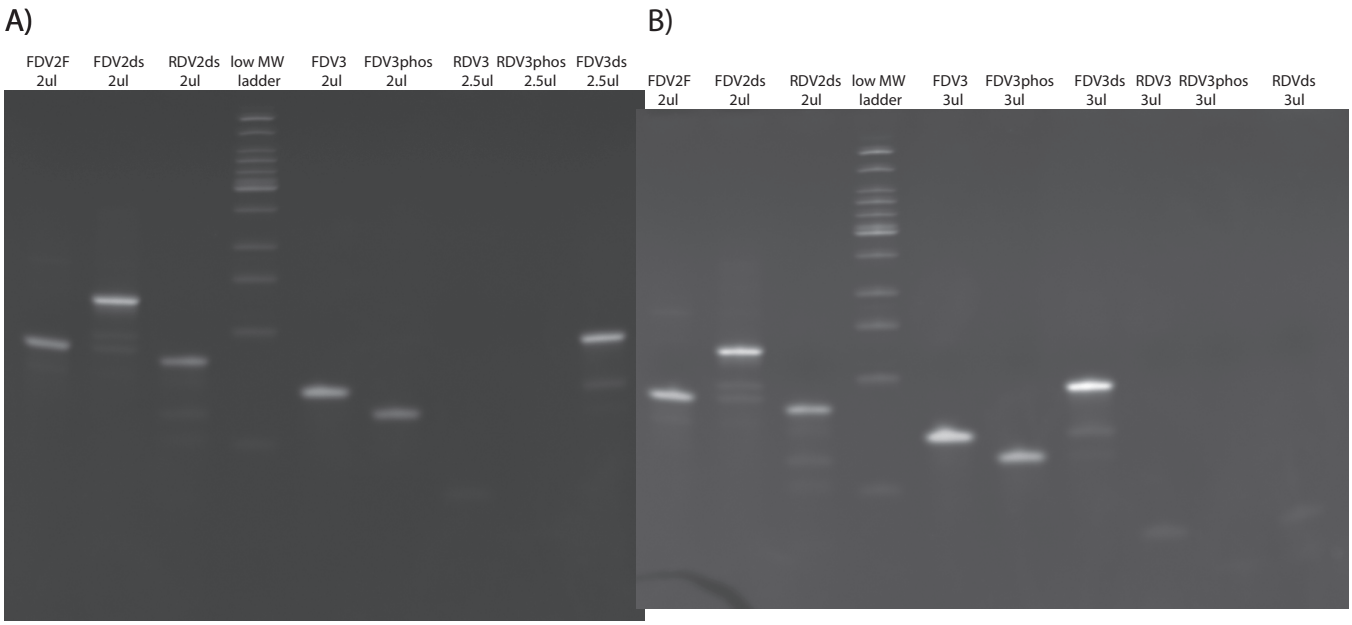


Figure 6.28: primers urea gel3

Brief Conclusions: The double stranded overhang piece migrates slower. The piece that migrates as single-stranded creates a second band (see in particular Figure 6.28b columns 5 and 7).

Ligating ends first attempt

Nov 17, 2006

I'm trying three reactions: 1) primers only 2) primers + 80mer, 3) primers + 120mer. I made 50 μM stocks of the 4 primers, and use 1.8 μl of each for the ligation (this is about 25x molar excess and corresponds to 1.8 μg of the large adaptor and 900 μg of the small one). I used 200 ng of the 80mer and 120mer blunt Phusion PCR products (4 μl).

Ligations were run for 12 hours at 16C followed by heat inactivation of the T4 ligase at 65C for 20 minutes.

Nov 19, 2006

I ran 5 μl (1/4) of the above ligation reactions on a 15% TBE Urea gel and post-stained the gel with EtBr (see Figure 6.29).

Brief Conclusions: Tue Nov 28 16:32:42 EST 2006

A look at the primer only lane shows that these primers are definitely ligating into longer pieces than they should (Figure 6.29). The primers have an overhang and are only phosphorylated on one end. So it shouldn't be possible to have long fragments and the possible bands should be much more continuous than I see here.

Next time run one lane with the same conc of primers and PCR, but without ligase, so I have something to compare to in the figure (see Figure 6.29). Above, I mentioned that the primers are

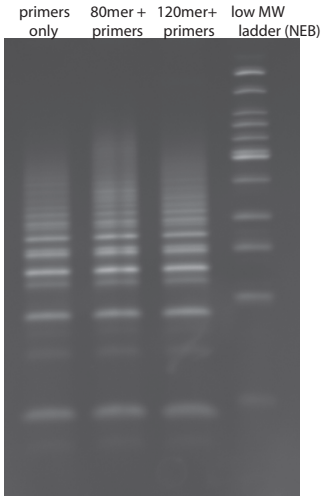


Figure 6.29: 12 hr ligations of excess adaptor primers to an 80-mer and 120-mer PCR products were run on a 15% TBE Urea gel for 50 minutes at 190V.

about 25x molar excess. This leads to a total primer pair molar excess of 50x. I think I need to lower this if I'm going to be able to see anything besides the primers on the gel. I also think raising the ligation temperature might help make things more stringent. I should try again with unphosphorylated primers like Shendure used. I should definitely try shorter ligation times.

Ligating ends second attempt

Nov 29, 2006

I lowered the primer concentration by 1/2 in the hopes of having clear results rather than a big smear of DNA. I also used 0.5 μl (200U), instead of the typical 400U. I varied the incubation times and used unphosphorylated primers (RDV2 and FDV2). Tested primer combinations and time variations are show in Figure 6.30.

Brief Conclusions: Mon Dec 4 17:56:12 EST 2006

This experiment was informative, unfortunately I'm not there yet. One thing to consider is that the 5' ends of the 80mer PCR product are NOT phosphorylated because the PCR primers aren't. The unphosphorylated priemrs like Shendure used do prevent self-ligation (see lane 3 and lane 10). However, they also don't seem able to ligate on their (at least not in the amount of time I used). On the other hand, I'd say the phosphorylated oligos are too efficient with this amount of ligase. I think 50-100U might be even better.

One thing that is wierd, the single phosphorylated dsDNA adaptors behave just as expected (lanes 1 and 2). They concatenate one time and make a second band. But when I stick the two adaptors into the same reaction (lanes 4 and 5), they ligate together in all kinds of different ways. I think I'll keep 15 minutes as the ligation time and try titrating the T4 units down.

To Do!!! make more 80mer, keep concentrations of the primers the same. use only 15 minute ligation time, but titrate down the amount of ligase (200, 100, 50). stay at 25C. try UREA and a TBE 6% (10μl on one 10μl on the other).

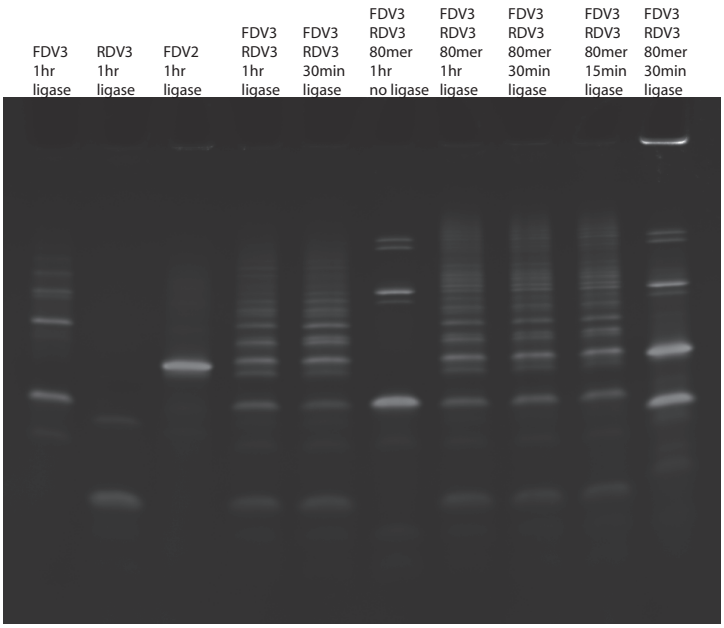


Figure 6.30: ligations of excess adaptor primers to an 80-mer run on a 15% TBE Urea gel for 50 minutes at 190V.

Ligating ends third attempt

Mon Dec 4, 2006

I made more of the 80mer[21] with 3, 100 μl PCR reactions. For each rxn I used 4 μl of 10 mM primer, 11 ng of plasmid, 45 μl H_2O , 50 μl Phusion mastermix; 30 sec 98C denature, then cycle: 5 sec 98C, 15 anneal, 15 extend at 72C. Ran 30 cycles, the first 5 were annealed at 60C, the last 25 were at 67C.

Tue Dec 5 20:31:55 EST 2006

I combined the 3 rxns into 2 tubes (150 μl each) and cleaned them up with a Qiagen PCR purification kit. To one of the two cleaned up rxns, I added 5 μl of T4 ligase buffer and 1 μl of T4 polynucleotide kinase (so that the blunt PCR producted would have 5' phosphorylations). I incubated at 37C for 30 min and cleaned up the rxn with a Qiagen PCR cleanup kit. The yields from all of these steps were:

Sample	DNA (ng/ul)	260/280	260/230	total yield
PCR sample 1	98.1	1.85	2.27	4.9 μg
PCR sample 2	90.8	1.89	2.27	4.0 μg
PCR sample 2 (after phosphorylation)	80.9	1.97	2.28	3.2 μg

I did my primer preparation a little differently this time. Instead of mixing all the stuff together, heating it up, cooling it, and adding ligase. I annealed all the primers separately in STE buffer (TE plus a 50 mM NaCl). I heated the adaptors up to 95C and dropped the temperature by 2C every 30 seconds until 4C.

[21] this is the same 80mer described in section 10.3.1 on page 397

I ran several different combinations of primers using different amounts of ligase (200, 50, or 10 Units per rxn). Half of the 20 μl rnxs were run on a 15% TBE UREA gel and a 6% TBE gel 6.31. Contrary to the previous attempt at these ligations, I ran this ligation at 16C.

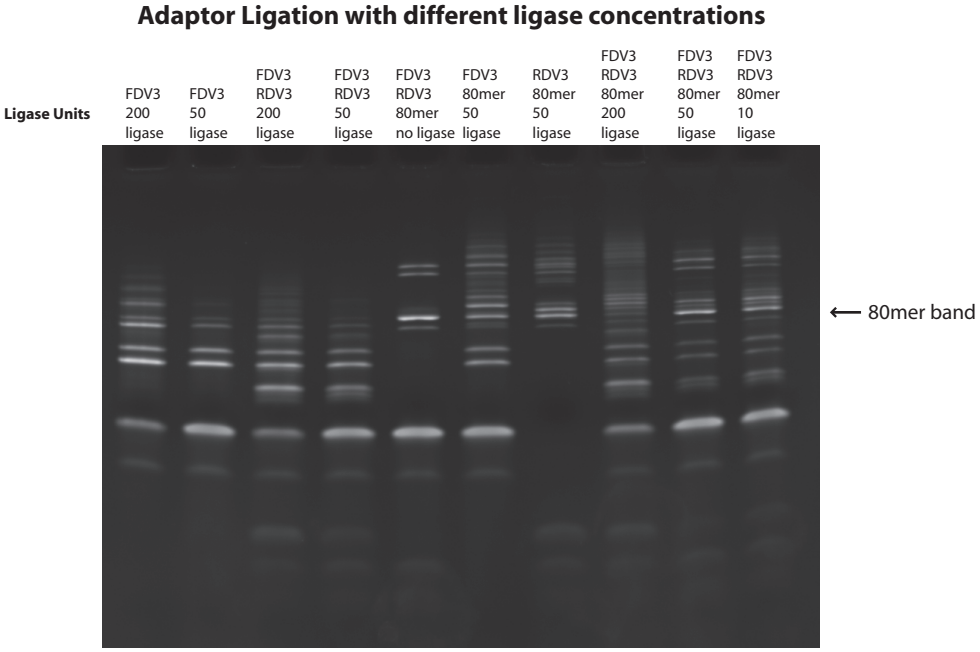


Figure 6.31: ligations of excess adaptor primers to an 80-mer run on a 15% TBE Urea gel for 50 minutes at 190V. I titrated the ligase concentration.

Brief Conclusions: Wed Dec 6 13:06:53 EST 2006

Houston we have progress! Finally, some insights that are leading to (hopefully) finishing this crap up. So the reduction of ligase concentration definitely helped out. The contrast between 200U and 50U is quite dramatic (see Figure 6.31 lanes 1 vs 2, 3 vs 4, 8 vs 9). This ligase unit reduction does little if anything to change the amount of the ligation products we're interested in (and when it does change them it makes more not less of the correct bands, because they don't end up in long concatamers). The reduction does lead to more of the unligated adaptors (this is to be expected and is a good thing). The 10U rxn was even better and is the first time that it is clear that different adaptors have been ligated to the 80mer PCR product. The 15% TBE Urea gel doesn't do the best job of showing this 6.31, because the separation isn't so good at the size. Would also be useful to have a DNA ladder to size confirm this stuff a little better.

Ligating ends fourth attempt

Dec 6, 2006

Based off the previous results (Figure 6.31), I diluted T4 ligase even further. Using 50, 10, 5, and 2 Units. This time I ran them on 6% TBE gels only (see Figure 6.33). I ran the first gel too long, so I ran a second one (Figure 6.34) and replaced lanes one and two with a 25bp and a 10bp ladder respectively.

Figure 6.32: ligations of excess adaptor primers to an 80-mer run on a 6% TBE Urea gel for 50 minutes at 200V. I titrated the ligase concentration. Unfortunately the gel went all crooked. For some reason it started out that only one side was moving. I didn't notice until 30 minutes later. I pushed the lid on tighter and it fixed the problem but the bands were already broken beyond repair.

Brief Conclusions: Mon Dec 11 15:52:52 EST 2006

I ran the first gel too long. However, I can still see the size-range I'm interested in. The number of bands is correct (Figures 6.33 and 6.34), though the correct bands are too faint. The sizes (in bp) of the bands are too high based off what I'd predict: 78=62, 82=92, 97=100, 112:113=117, 128=130, 144=152. The second gel (Figure 6.34) leads me to believe that maybe I made the run PCR product as the band is at 120 bp not at 80 bp. Despite the differences in the absolute bp estimates from versadoc, the relative difference in base pairs is similar in the gel to what is expected (expect:gel1:gel2, 15:18:14, 6:8:4, 15:17:15, 16:18:16). I have no idea what that big band is. Based on gel 2, I'd guess it is the concatamerized PCR product. But for gel 1, the numbers don't work out right. Notice that the band that corresponds to two different possibilities is the brightest concatamer band (the number in both figures is in italics). This band correponds to FDV3:PCR and RDV3:PCR:RDV3.

I'm pretty much there, the most important thing is to get that band a little brighter (and thereby making the PCR product band a little less strong). I think I need to increase the concentration of each primer, try 50 and 100 U of T4 ligase and ligate for 30 minutes instead of 15. Even as it is now though at least I can see the correct band.

6.7.3 Ligating ends using an RCA product *again*

Mon Dec 11 19:06:17 EST 2006

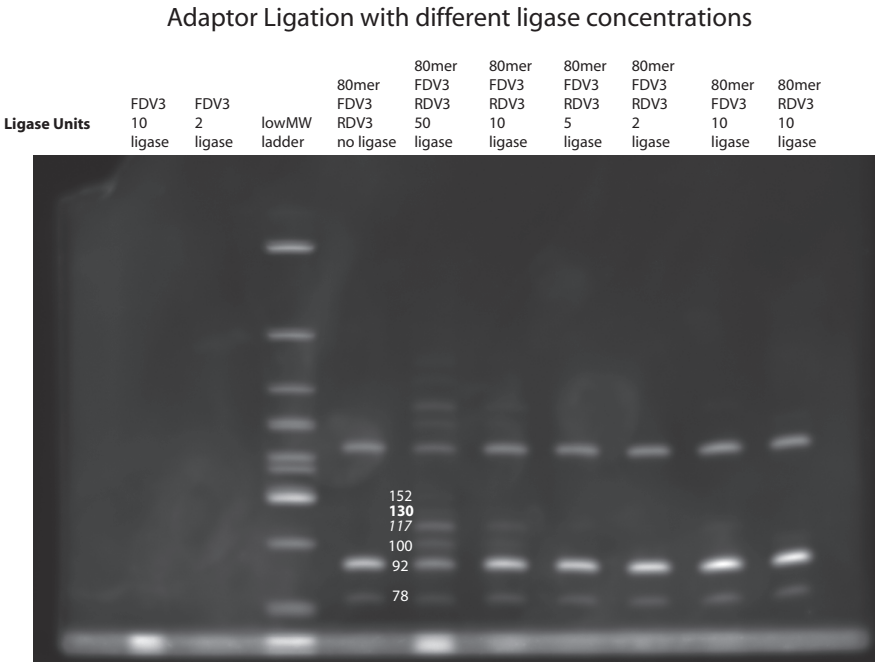


Figure 6.33: ligations of excess adaptor primers to an 80-mer run on a 6% TBE Urea gel for 45 minutes at 200V. I titrated the ligase concentration. Unfortunately, I ran the gel for too long. The band lengths I would be interested in are 30, 62, 82, 97, 112, 113, (128), 144.

I'm starting to get the hang of things with the PCR product. Now I want to go back to the RCA product and see if I can get things working with the MmeI digested, end-repaired tag.

A modification of my earlier approach I'm leaning towards the idea of not size-selecting before I end-repair and ligate on the two different adaptors. I feel like I lose too much and make the DNA too junking by selecting the thing from the polyacrylamide gel. Instead I'm going to use a Qiagen PCR cleanup kit to remove the large undigested RCA stuff (which I know from previous experience it does [22]) and the small nucleotide stuff that would use up the enzymes in the end-repair kit and lead to ligation junk problems in the adaptor ligation steps.

Then I'll ligate my adaptors to all of the Qiagen cleaned up RCA products. This should result in a smear with 5 bands (82, 97, 112.5 [big band], 128 [band we want], 144). I might also try use the dual-biotin oligo to capture and further enrich the 128mer tag prior to running the 6% gel (might need to be UREA if I use the biotin-capture) and size-selecting the correct band for PCR amplification.

testing the Qiagen cleanup and biotin oligo selection procedures

Mon Dec 11 19:21:48 EST 2006

I'm digested 15 μl (10 μg) of RCA product (the one from page 270) in 5 μl NEB4, 5 μl MmeI, 2 μl SAM (diluted 1/20 from the 32 mM stock), 23 μl H_2O . The digestion was for 30 min 37C followed by heat deactivation for 10 minutes at 65C.

[22] see sections 6.6.2 and 6.6.2 starting on page 269

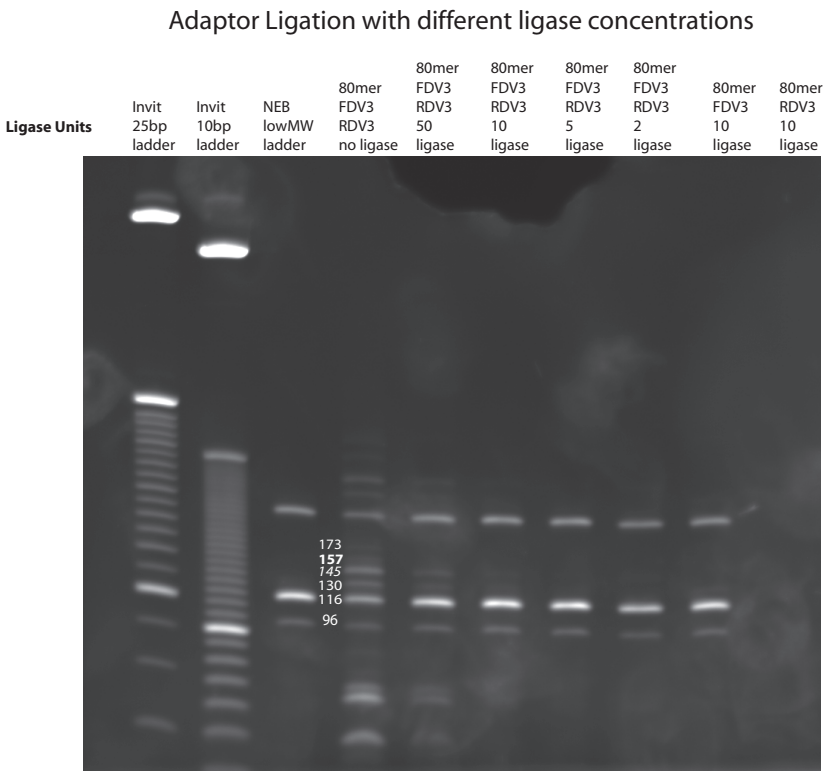


Figure 6.34: ligations of excess adaptor primers to an 80-mer run on a 6% TBE Urea gel for 45 minutes at 200V. I titrated the ligase concentration. Unfortunately, I ran the gel for too long. The band lengths I would be interested in are 30, 62, 82, 97, 112, 113, (128), 144. This gel is the same as Figure 6.33 except that the first to lanes were replaced with the 25bp and 10bp ladders from Invitrogen, and the gel wasn't run so long.

I left 12.5 μl and cleaned up the remaining 37.5 μl with a Qiagen PCR cleanup kit to remove the very large and very small DNA. I eluted into 31 μl and spec'd the resulting DNA on the nanodrop:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**total yield**
digested RCA after cleanup	56.7	1.84	2.27	1.76 μg

Biotin-oligo DNA capture: first try Based on a recommendation from Natalia Broude, I'm not going to use the recommended Dynal salt concentration of 1M[23]. I'm using a Binding and Wash buffer (BW) of 100 mM final NaCl concentration (10 fold lower than recommended). Binding and wash buffer is just TE with salt added.

I'm going to use the following protocol:

[23]she claims this is way to high and you can't ever get your captured oligo off

1. wash 5 μl of Dynal beads in 5 μl of 2x BW buffer
2. repeat step 1
3. resuspend beads in 12.5 μl of 2x BW buffer
4. add 12.5 μl of oligo *plus* the dual-biotin 30mer (this will be 1 μl of the 50 μM working stock) [24].
5. heat the 25 μl soln to 98C for 2 minutes
6. flick tube
7. incubate in 50C for 1 hr to binding oligos; give tube a few flicks every 15 minutes or so
8. wash 2x in 1x BW buffer
9. resuspend in 25 μl of 0.1N NaOH at 50C for 5 minutes to remove captured oligo
10. place on magnetic stand one minute to capture beads; keep the supernatant (the solution containing the now freed oligo)
11. add 25 μl of 1M Tris (to adjust the pH back)
12. add 425 μl of TE (to make the near the maximum volumn for the microcon column)
13. concentrate with YM30 microcon; spin 12min at 14000g [25]

I loaded the uncleaned up RCA, the Qiagen cleaned RCA, the oligo selected RCA and a NEB lowMW ladder and a Invitrogen 25 bp ladder (0.5 μl = 0.5 μg) onto a 6% TBE polyacrylmide gel. The original Qiagen cleaned DNA floated right out of the well (Figure 6.35, maybe it has something to do with EB buffer?). So I added 2.5 μl TE and 3 μl of TBE Hi Density buffer to the last remaining part of the Qiagen cleaned sample and loaded it into the final lane.

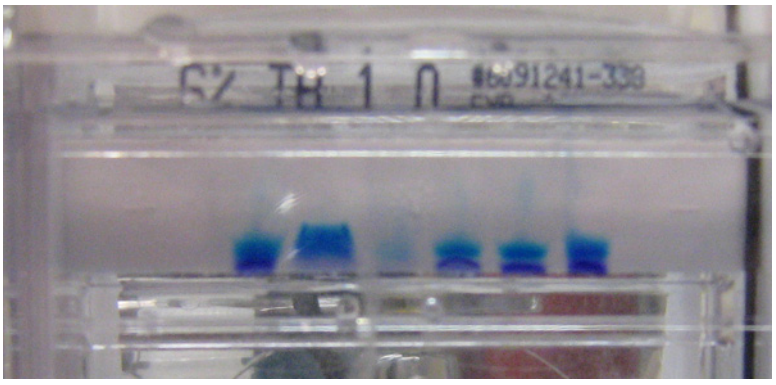


Figure 6.35: The Qiagen EB buffer plus TBE Hi Density buffer would not settle into the bottom of the well and diffused out.

I ran the gel for 35 minutes at 200V. Stained 20 min in EtBr and 10 min destain in H_2O (Figure 6.36).

[24] I used 567 ng of the RCA product
[25] had to add 20 μl TE to the dried membrane to elute

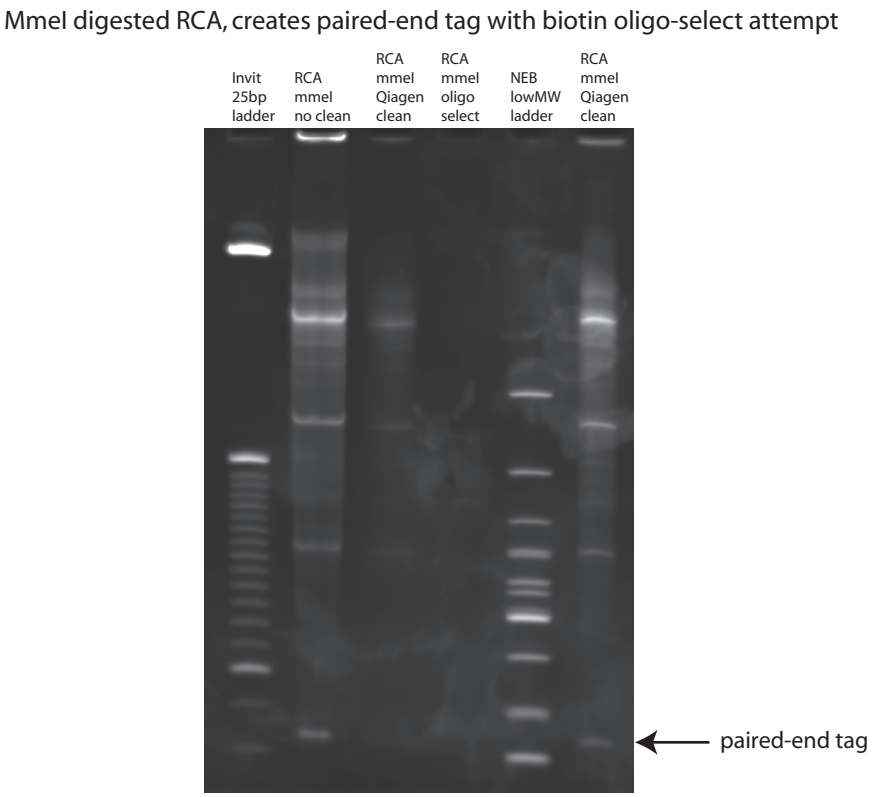


Figure 6.36:

Brief Conclusions: Wed Dec 13 19:19:08 EST 2006

The float away lane was faint as expected (Figure 6.36 lane 3). But lanes 2 (uncleaned digestion) and lanes 6 give the cleanest view so far of our PET. Running one-fourth of one digestion seems a lot clearer than previous fuzzy bands on hard to interpret gels as occured in my previous attempts of loading the entire digestion on one lane. Also using EtBr was a big benefit over Sybr Gold with the versadoc (see Figures 6.22, 6.21, and 6.20 for previous gels with the PET tags).

The biotinylated oligo selection either didn't work or didn't use enough DNA so that not enough DNA was recovered to make a visible band (Figure 6.36 lane 4). Overall, I think things look pretty promising for adding the adaptors to the RCA products.

try to add adaptors Dec 13, 2006

The above experiment showed me that the PET tag is fairly clean when I don't have too much DNA (Figure 6.36). Now I want to see if I can get the adaptors on there. Maybe in previous attempts I just had too much DNA?

I'm going to do the same digestion: 15 μl (10 μg) of RCA product (the one from page 270) in 5 μl NEB4, 5 μl MmeI, 2 μl SAM (diluted 1/20 from the 32 mM stock), 23 μl H_2O . Digest for 30 min at 37C followed by heat deactivation for 10 minutes at 65C. After cleaning the rxns with a Qiagen PCR cleanup kit the yields were:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**total yield**
digested RCA after cleanup	63.3	1.91	2.20	2.2 μg

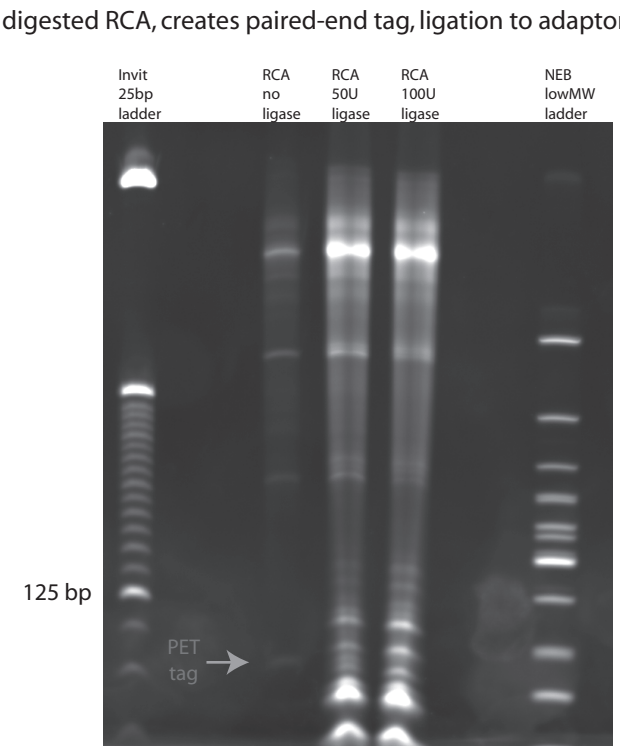


Figure 6.37: The tag band is appx 82 bp. After that we expect to see: 97, 112:113, 128, 144.

Brief Conclusions: Thu Dec 14 23:01:12 EST 2006

I can't seem to get a good polyacryamide gel any more. They've always been finicky but this is ridiculous. I don't know how well this worked because the gel sucks (Figure 6.37). It doesn't look horrible (the rxn not the gel); it looks like I had too much adaptor primer.

To Do!!! use less DNA (split into 3), use less primers, clean DNA after ligation, keep same amount of ligase, run 2 two parallel ligations tomorrow, 1) use polyacrylamide, 2) use TAE and Nusieve; run for a LONG time

try to add adaptors *again* Thur Dec 14, 2006

Using the same protocol as yesterday, but today I'm running two digestions, one for a polyacrylamide gel and one for a Nusieve gel.

RCA digestion yields were:

Sample	DNA (ng/ul)	260/280	260/230	total yield
digested RCA sample 1 after cleanup	78.0	1.91	2.18	2.7 μg
digested RCA sample 2 after cleanup	70.9	1.89	2.20	2.5 μg

After end repair and clean up of the end repair with a Qiagen PCR kit, yields were:

Sample	DNA (ng/ul)	260/280	260/230	total yield
end-repaired PET sample 1 after cleanup	71.9	1.96	2.30	2.2 μg
end-repaired PET sample 2 after cleanup	71.5	2.02	2.36	2.2 μg

For the ligations I used 1/3 of the cleaned up sample 1 from the table above for each of the two reactions (2/3 total). I also ran 1/6 of the unligated RCA digest on a polyacrylamide gel with the two ligations. I cleaned the 20 μl ligations with a Qiagen PCR kit and eluted into 30 μl . I ran 20 μl (because the entire 30 μl wouldn't fit) on a polyacrylamide gel (Figure 6.38).

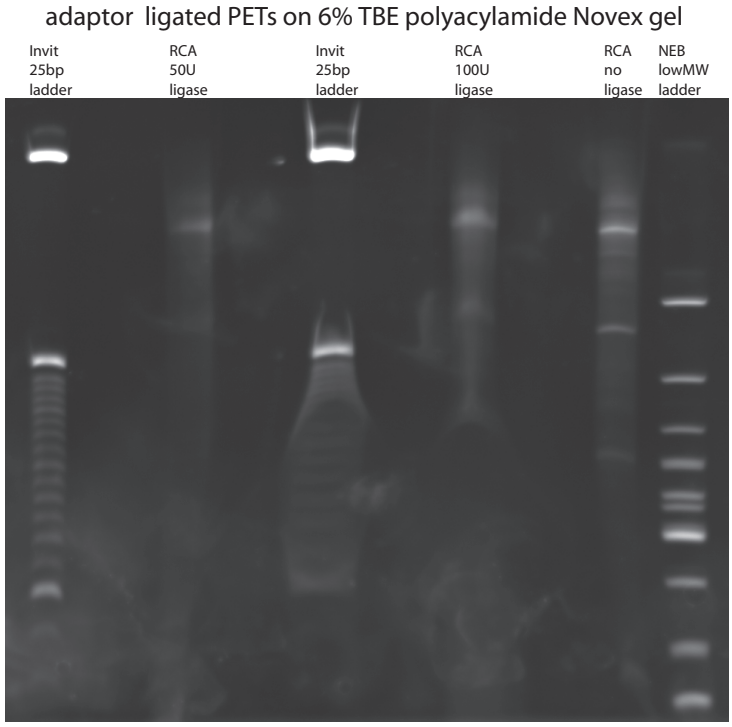


Figure 6.38: 20 μl of the 30 μl Qiagen cleaned ligation reactions were run on a 6% polyacrylamide gel.

Once again having problems with consistent running of the Novex precast gels, I decided to run the remaining 10 μl of each ligation and the remaining 1/6 of digested RCA onto a 3.5% Nusieve TBE gel Figure 6.39.

Brief Conclusions: Well, there are some nice conclusions that we can make here: 1) Novex polyacrylamide gels either suck or I suck at using them (Figure 6.38); 2) Nusieve is *much* better (Figure 6.39); 3) Nusieve requires much more DNA per lane (compare the 25 bp ladder between the two figures); 4) the exACTGene 50 bp ladder is a beautiful addition because it has a band at 100 and at 112 bp which is very similar to two of the bands I'm really interested in (because if I can find those two then I know that the band I want comes next); 5) add the 25 bp ladder and you have pretty much a ladder band corresponding to every adaptored PET I am interested in (see my bp annotations on the side of the bk subtract gel). 6) If you look really close (and I stress the *really* in the background subtracted image I think the adaptored PET band of interest might be faintly visible

Brief Update *Tue Dec 19 15:11:48 EST 2006*: When I ran the Nusieve gel (Figure 6.39) the mAmps was very high and I had trouble reaching a decent voltage. The next time I ran a Nusieve gel, I used 0.5x TBE and the mAmps were much lower and it seemed like the gel didn't get as hot.

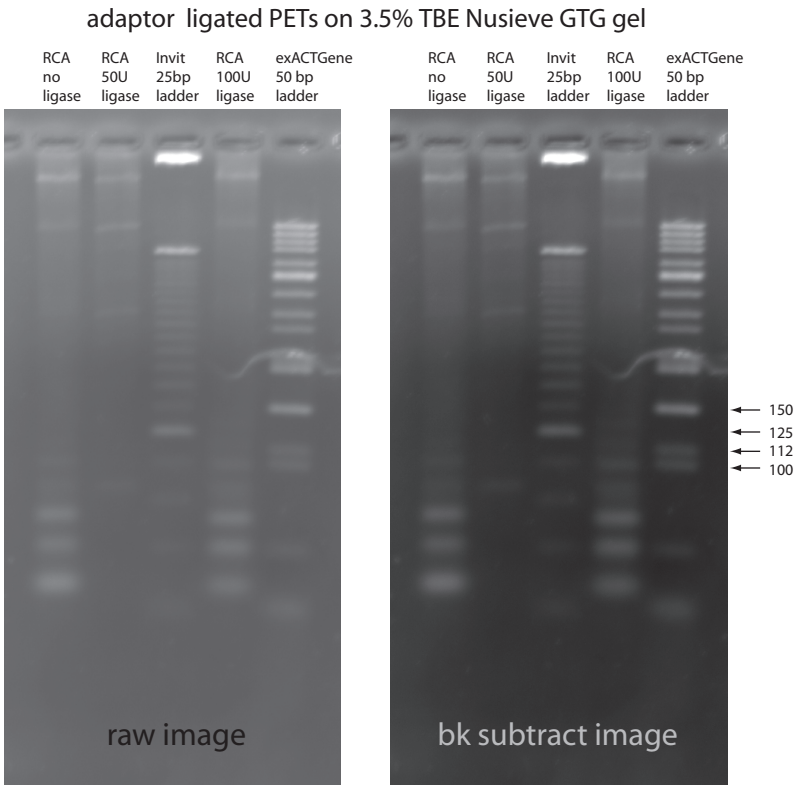


Figure 6.39: 10 μl of the 30 μl Qiagen cleaned ligation reactions (the left overs from Figure 6.38) were run on a 3.5% Nusieve TBE. The image on the right has been background subtracted using the Versadoc software.

try to add adaptors *again*, *again*; this time we'll cut the gel slice I used sample 2 from the above RCA digestion. I ligated it in the same way as above and ran 20 μl on a Nusieve gel (previous Nusieve gel only had 10 μl on it).

I cut out the band for 50U and 100U at approximately 128 bp.

Brief Conclusions: Things are looking up since I switched to Nusieve (Figure 6.40). I've heard heard MetaPhor has even better resolution but is very fragile and not low-melt (so you have to do some sort of gel cleanup). I think I still need to run more DNA. I also need to use more DNA in the ligation. Last, I would forgo the EtBr and just use sybr gold and the crappy imaging system. I post-stained the gel with SybrGold after post-staining it with EtBr for imaging. Sybr gold isn't nearly as bright if EtBr was on the DNA first. I cut out the bands at the correct sizes. The bands were pretty dang faint though, so hopefully next time I have a slightly more obvious band to cut out.

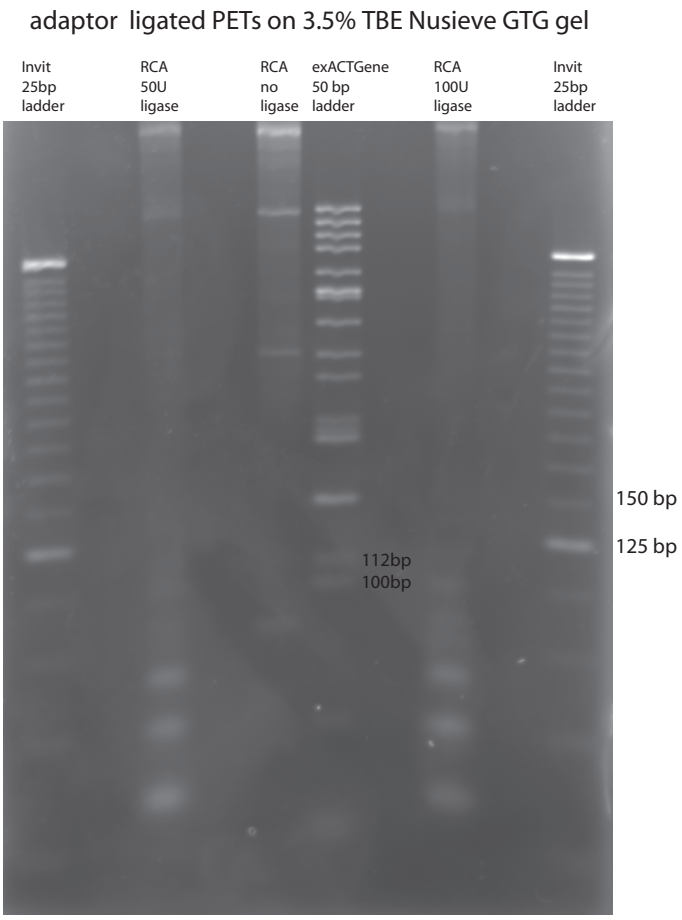


Figure 6.40: 20 μl of the 30 μl Qiagen cleaned ligation reactions were run on a 3.5% Nusieve TBE. The appx 128 bp band was extracted.

6.8 Amplifying the adaptored tags

6.8.1 PCR Primers

I want to make a variety of primers for amplifying the adaptored DNA so I can clone it or do what ever else I like. Below are the primers I'm using for now

::

```
Primers from Shendure et al:

Unmodified forward primer
5' CCTCTCTATGGGCAGTCGGTGAT

Reverse primer
5' CTGCCCCGGGTTCCTCATTCTCT

------------------------------------------------
```

```
Forward primer for directional cloning EcoRI
5' TAGAATTCCCTCTCTATGGGCAGTCGGTGAT


Reverse primer for directional cloning BamHI
5' ATGGATCCCTGCCCCGGGTTCCTCATTCTCT


------------------------------------------------
Forward primer for directional cloning USER
5' CCTCTCTATGGGCAGTCGGTGAT

Reverse primer for directional cloning USER
5' CTGCCCCGGGTTCCTCATTCTCT
```

PCR amplification of cut adaptored tag Mon Dec 18, 2006

I PCR amplified the band taken from 6.7.3 on page 291 (Figure 6.40). I heated the gel slice up to 65C for appx 20 minutes and used 5 μl in a 50 μl PCR with Phusion Taq. I ran one reaction with each of the two gel slices taken and one rxn with each of the primers sets 1) EcoRI/BamHI added primers AND 2) blunt primers. The blunt primers used an annealing temperature of 63C for all cycles. The sticky end primers used 63C for the first 5 rounds and 66C for the remaining 25 rounds.

Used 10 μM primer, melt 10 sec 98C, extend 15 sec 72C. PCR rxns were run out on a gel 6.41.

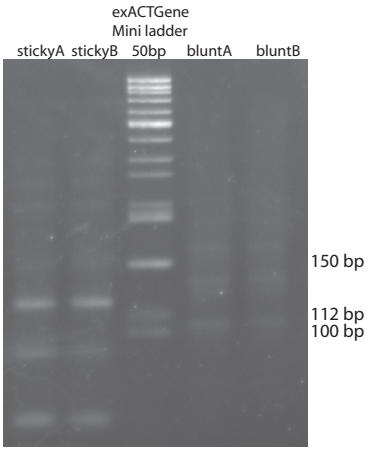


Figure 6.41: 5 μl of Nusieve GTG agarose gel slice (128bp) was amplified for 30 cycles in a 50 μl rxn.

Brief Conclusions: It appears that a band of approximately the correct 128 bp size is visible. But several other bands are present as well.

PCR amplification of cut adaptored tag: try 2 Thu Jan 4, 2007

I repeated the PCR from the gel as I did on Dec 18. I wanted to try and determine why the blunt rxn didn't work. I ran everything like above except I didn't use the two step proceedure with the sticky ends. I keep all the rxns at 63C for the entire time. The PCR was run out on a 2% SB gel 6.42.

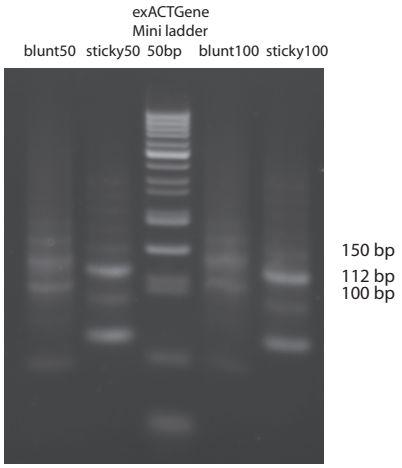


Figure 6.42: 5 μl of Nusieve GTG agarose gel slice (128bp) was amplified for 30 cycles in a 50 μl rxn.

Brief Conclusions: I looks like the blunt primers consistently fail (Figures 6.41 and 6.42.

6.9 Cloning the adaptored tags

Now that it looks like I might have the correctly adaptored PET tag band cut and amplified, I'm gonna try to clone and sequence it.

6.9.1 cloning the sticky PCR product into pUC19

Fri Jan 5, 2007

The sticky PCR was the only one to work consistently (Figures 6.41 and 6.42), so that's what I'm gonna try to clone into pUC19.

I cleaned the PCR reactions with a gel clean up kit, because the 5 μl in the PCR made the PCR rxn pretty gelly. In retrospect, I don't know why I didn't cut the band out since there were clearly multiple bands in the PCR? The specs of these clean ups looked pretty dirty. The yields from the clean up of all four rxns were:

Sample	DNA (ng/ul)	260/280	260/230	total yield
blunt 50	44.2	1.88	0.39	1.33 μg
sticky 50	42.5	1.87	0.19	1.28 μg
blunt 100	38.5	1.83	0.51	1.16 μg
sticky 100	104.0	1.64	0.64	3.12 μg

I cut 2 μg of pUC19 with EcoRI and BamHI and cleaned it up with a Qiagen PCR cleanup kit. I cut all of the purified sticky 50 and sticky 100 PCR products with EcoRI and BamHI and cleaned them in the same manner.

I ligated the sticky 50 and sticky 100 in two different rxns consisting of: 1 μl T4 ligase, 2 μl T4 ligase buffer, 8 μl cut PCR, 2 μl cut PCR (I meant to use 1 μl but screwed up), 8 μl H_2O . 2 μl of each ligation was tranformed into DH5alpha compentent cells.

6.9.2 checking the clones

Mon Jan 8, 2007

I picked five sticky-50 clones (samples 2a-e) and 6 sticky-100 clones (samples 4a-f). I PCR amplified them with the M13 primers. The primers should add 80-90 bp, so the final piece should be around 200 bp. I used 1 μl of miniprep in each rxn. The first nine rxns were run on an agarose gel (Figure 6.45).

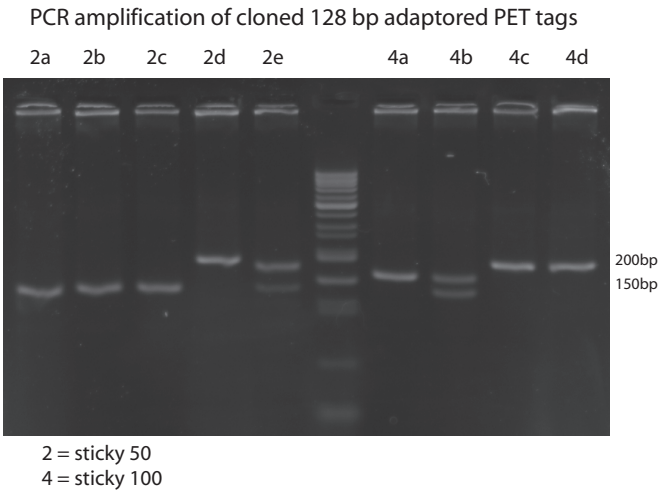


Figure 6.43: Miniprepped pUC19 cloned 128bp sticky adaptored PET tags were amplified with M13 primers

Tue Jan 9, 2007

I spec'd five of the miniprepped samples to send them to agencourt for sequencing: 2b, 2c, 2d, 4a, 4b.

Sample	DNA (ng/ul)	260/280	260/230	total yield
tag 2b (2)	235.2	1.90	2.17	11.8 μg
tag 2c (3)	190.3	1.91	2.14	9.51 μg
tag 2d (4)	272.0	1.92	2.24	13.6 μg
tag 4a (7)	104.0	1.91	2.27	10.2 μg
tag 4b (8)	104.0	1.92	2.26	12.4 μg

600-1000 ng of each of the five samples was sent to agencourt for sequencing using the M13F primer.

Brief Conclusions: Didn't work, no tags were in the pUC19 vector.

6.10 Cloning the unadaptored tags

Many moons ago, when this project was going so well, I skipped sequencing the unadaptored PET tags because I thought it was a waste of time, given that everything looked correct. Not that nothing works, and I haven't made forward progress since that day, I'm taking a step back to try and figure out what went wrong.

6.10.1 blunt coloning unadaptored tags try 1

Tue Jan 30, 2007

I cut the tags from RCA2b (15 μl = 10 μg), ran them on a gel 6.44, and cloned them into an Invitrogen Zero Blunt PCR for Sequencing kit. I sent five of the clones out for sequencing today. Although now that I look at the length of the inserts in detail, unfortunately, it looks like the PCR check of insert verification is NOT long enough.

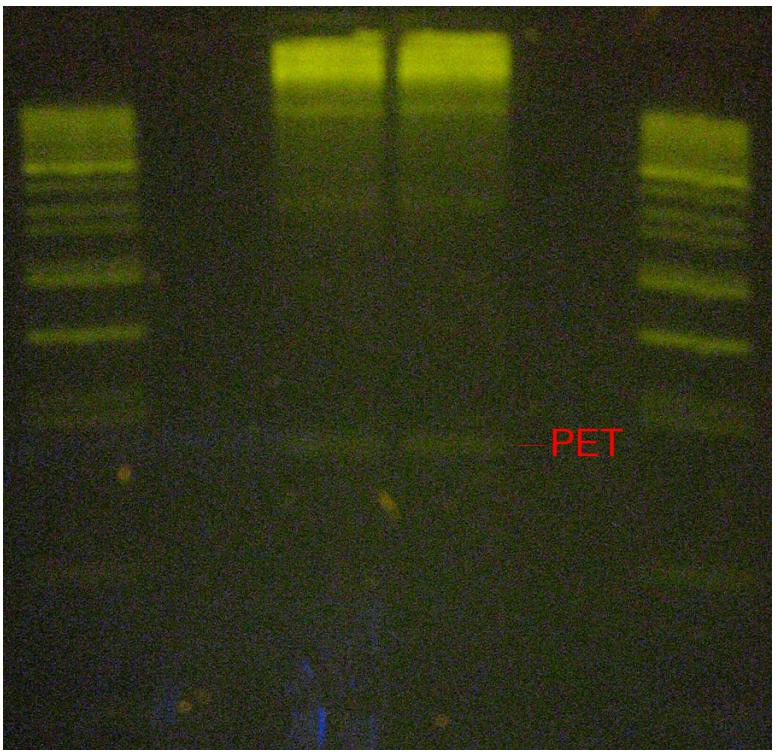


Figure 6.44: RCA2b tag was cut from lanes 3 and 4 and combined into one Qiagen gel cleanup

The yields for the minipreps were (sequenced ones are in italics)

Sample	DNA (ng/ul)	260/280	260/230	total yield
(1)	327.3	1.89	2.10	16.4 μg
(3)	68.8	1.95	2.17	3.4 μg
(4)	133.9	1.83	2.12	6.7 μg
(5)	328.0	1.88	2.17	16.4 μg
(6)	539.7	1.89	2.20	27 μg
(7)	493.5	1.87	2.21	24.7 μg

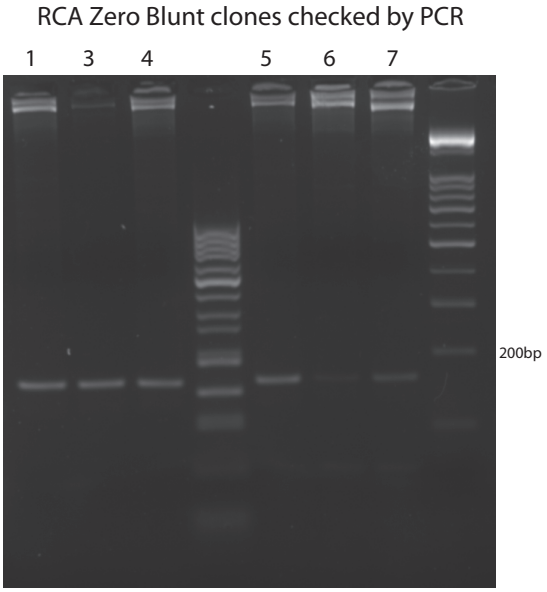


Figure 6.45: Miniprepped Zero Blunt PETs, inserts don't seem to be long enough (PCR minus insert should 170bp, with insert should be 250 bp)

I'm trying again with more RCA DNA. This is the last drop I have of RCA DNA, so hopefully this works...

Brief Conclusions: The sequences came back with no inserts. Technically, having the same vector close on itself is impossible with the Zeroblunt kit, because that would result in there being expression of the lethal ccdB toxin. However, in all the sequences I got back there was *no* insert, but rather one or two base-pairs had been added/removed[26] from the vector resulting in a ccdB gene that was frame-shifted and therefore not a properly folded lethal protein.

6.10.2 blunt coloning unadaptored tags try 2

Wed Jan 31 17:24:57 EST 2007

I'm trying again, but this time I'm digesting more of the RCA. Actually, I'm digesting all of the remaining RCA, so hopefully this works! I only had about 30 μl (20 μg) of the RCA. Which I digested with 15 μl of MmeI at 37C for 15min and heat deactivated at 65C for 10 minutes. Instead of a Qiagen Cleanup, I did an EtOH precipitation, and eluted into 15 μl of TE, added 2.5 μl loading dye and ran it all in one lane of a 3.5% TBE gel with Sybr Safe.

Thur Feb 1, 2007

I cut the gel on the transilluminator (at jpeg picture here) and cleaned it up with a Qiagen gel clean up kit, eluting into 34 μl of EB buffer. I then repaired the ends with a *new* End-it End repair kit (note: I should throw away the old one, it's been freeze-thawed too many times). I increased the volume to 400 μl with TE cleaned up the end-it reaction with Phenol/Chloroform and then with Phenol. Then I concentrated the DNA with EtOH precipitation and eluted into 15 μl of TE.

[26] mostly there were small deletions

I used 4μl of 15 μl of end-repared in the ZeroBlunt reaction [Invitrogen].

Fri Feb 2, 2007

I picked eight colonies from the ZeroBlunt cloning.

Sat Feb 3, 2007

5 of the 8 colonies grew in kanamycin (the plates I used for the cloning were amp, but this vector has two resistance genes). I minipreped the 5 and did a PCR with the M13 primer set.

Mon Feb 5, 2007

Ran the PCRs on a gel (Figure 6.46). Same junk with the ZeroBlunt, looks like no good inserts.

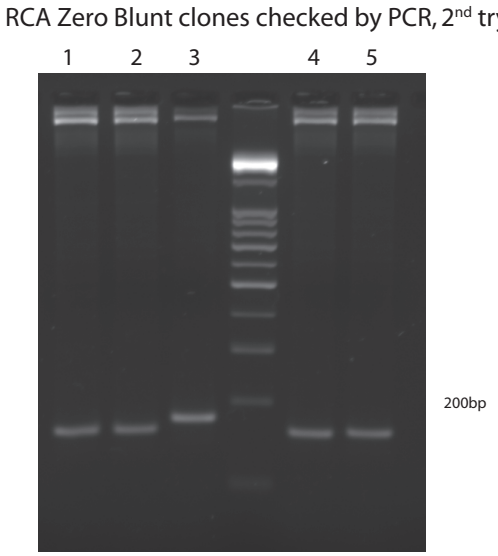


Figure 6.46: PET tags were end-repaired and cloned into the ZeroBlunt kit for the secone time. It looks like once again the insert failed to go in there. (See Figure 6.45 for the other failed attempt)

Brief Conclusions: I've lost a little faith in the folks at Invitrogen and their ZeroBlunt kit. Clearly I've got an appx 85 bp piece that should be blunt that just won't go into there superdupper vector.

6.10.3 blunt coloning unadaptored tags try 3, going old-school

Mon Feb 5, 2007

Clearly the ZeroBlunt kit and I aren't getting along. I'm going to try this the old fashion way. I digested 2μg of pUC19 with SmaI for 30 min at 25C and 20 min at 65C (to deactivate) to make a blunt vector. I cleaned the rxn with a Qiagen PCR cleanup and eluted into 30 μl of EB. I took 10 μl of the cleaned up digestion and aded 1.2 μl Antarctic phosphatase buffer and 1 μl antarctic phosphatase. I incubated this 30 min at 37C and 5 min at 65C.

For the ligation, I used 1 μl of the dephosphorylated vector, 4 μl of the end-repaired, gel selected PET tags (this was from the 11 μl remaining after I used 4μl in the previous experiment with the ZeroBlunt kit). I used 2 μl of T4 ligase buffer, 12 μl H_2O , and 1 μl *high concentration* T4 DNA ligase [NEB]. I ran the ligation for 10 min at RT and deactivated for 20 min at 65C.

I transformed 2 μl into DH5alpha, incubated on ice 5 minutes, heat shock 30 sec at 42C, back on ice, add 250 μl LB. Shake 200 rpm for 1hr at 37C. Plated 100 μl .

Tue Feb 6, 2007

Only had 3 white colonies, so I picked them all. I also replated the remaining ligation, so I can have more colonies.

Wed Feb 7, 2007

All 3 grew and I miniprepped, PCR'd and ran them on a gel (Figure 6.47). With the replate from yesterday, unfortunately I forgot to add Xgal!!!!!! :(So I initially thought, I'd just have to pick them and hope I had inserts. However, I talked Jamey for a while and he had many ideas from copying the plate with a piece of nitrocellulose membrane to picking them all in replating them with a stamper on a new plate (note the replate had probably around 120 colonies so at least that was good). However, I tried out his riskiest idea. I lifted the gel with a sterilized spatula and squirted around 300 μl of sterile H_2O mixed with Xgal. I did this 4 times, breaking the plate into four quarters. For each quarter, I used the amount of Xgal you'd normally use for one gel (so in the end I used 4x the normal amount). The idea (which I pretested on a test gel) was that the liquid would diffuse up to the colonies and allow them to change color without the problem of colony mixing that would occur of you added the liquid to the top of the plate (which would probably screw everything up). I hacked plate in the 37C incubator for about and hour and the blue-white colonies were clear as day. Very nice recovery of a big-screwup, thanks to Jamey!!! I picked 8 colonies and grew them overnight in LB.

Bio-cheats: If you forget to add Xgal to a blue-white cloning plate, the next morning (after you recover from your self-anger at your own stupidity for almost wrecking your sample) dilute your normal Xgal amount into 1.2 ml of H_2O . Mentally partion the agarose plate into 4 equal parts and for each part, leift the gel with a steriled spatula, add 300 μl of the Xgal H_2O mix, and let the gel back down. Put the plate in the proper incubator for 30-60 minutes and you should see your blue-white colonies just like if you'd done it right in the first place.

Thur Feb 8, 2007

I ran 8 minipreps and PCRs from the picked colonies. I didn't have enough time to run them on a gel.

Fri Feb 8, 2007

I didn't have time to check the insert lengths on the replated colonies before the agencourt courier got to BU. I spec'd all 11 samples and sent the 3 original pUC19-PET samples plus two of the replated ones. Yields for all of the minipreps are (r indicates from the replate; those sent out for sequencing are in italics):

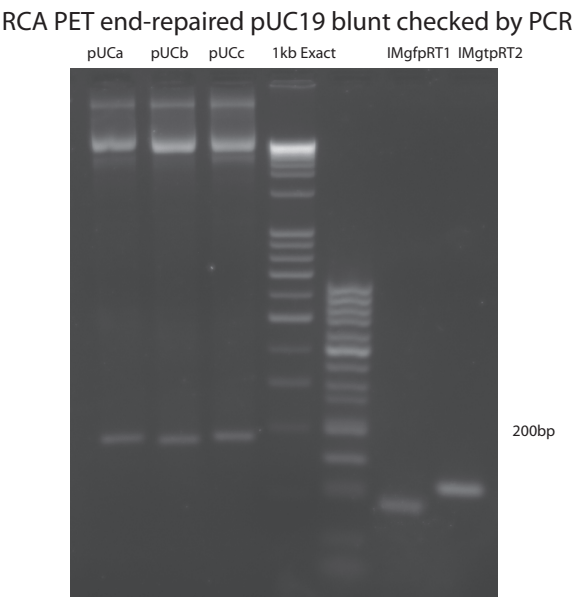


Figure 6.47: PET tags were end-repaired and cloned into dephosphorylated pUC19 at the smaI site (first three lanes). It appears the inserts are the proper size given the size of the insert (85bp) and the size PCR band given the additional stuff amplified outside this 85 bp by the M13 primers. The additional two lanes were just (successful) tests of two sets of RT-gtp primers for use in a different project.

Sample	DNA (ng/ul)	260/280	260/230	total yield
a	389.8	1.90	2.25	19.5 μg
b	511.8	1.89	2.19	25.6 μg
c	387.3	1.88	2.25	19.3 μg
r.a	195.3	1.87	2.09	9.8 μg
r.b	65.0	1.86	2.53	3.3 μg
r.c	144.6	1.77	2.26	7.2 μg
r.d	208	1.87	2.07	10.4 μg
r.e	196.3	1.88	2.07	9.8 μg
r.f	191.8	1.88	2.29	9.6 μg
r.g	304.7	1.87	2.25	15.2 μg
r.h	191.1	1.93	2.20	9.6 μg

```
-----   SEQUENCE tag a   -----

 --- TAG_FRONT ---
(4034053..4034070) rrsA, 16S ribosomal RNA (rrsA)

 Score = 36.2 bits (18), Expect = 4e-04
 Identities = 18/18 (100%)
 Strand = Plus / Minus

 Query: 1       cacggagttagccggtgc 18
                 ||||||||||||||||||
 Sbjct: 4034070 cacggagttagccggtgc 4034053
```

```
 --- TAG_BACK ---
 (4033583..4033594) rrsA, 16S ribosomal RNA (rrsA)

 Score = 24.3 bits (12), Expect = 1.4
 Identities = 12/12 (100%)
 Strand = Plus / Plus

 Query: 7       tgaacgctggcg 18
                 ||||||||||||
 Sbjct: 4033583 tgaacgctggcg 4033594

-----   SEQUENCE tag b   -----
no match (only looked by hand, very much not thorough)

-----   SEQUENCE tag c (awfully short)   -----
(4208234..4208252) rrlE, rrlE 23S ribosomal RNA

 Score = 38.2 bits (19), Expect = 1e-04
 Identities = 19/19 (100%)
 Strand = Plus / Minus

 Query: 1       cccggttcgcctcattaac 19
                 ||||||||||||||||||
 Sbjct: 4208252 cccggttcgcctcattaac 4208234

(4208100..4208117) rrlE, rrlE 23S ribosomal RNA

 Score = 36.2 bits (18), Expect = 4e-04
 Identities = 18/18 (100%)
 Strand = Plus / Plus


 Query: 1       ggcagtcagaggcgatga 18
                 ||||||||||||||||||
 Sbjct: 4208100 ggcagtcagaggcgatga 4208117

-----   SEQUENCE tag r.a   -----
no match (only looked by hand, very much not thorough)

-----   SEQUENCE tag r.b   -----
couldn't uniquely map tag
```

Raw data in Word format.

Brief Conclusions: Finally got that stupid little tag cloned into a vector (Figure 6.47)! Unfortunately, I have less than the agencourt minimum of 5 sequences.

6.11 New circularization adaptors

```
circularization adaptor 2  5' p-ATCGCA GCATCG ACG
                                    CGT CGTAGC TGCCTA-p 5'

circularization adaptor 3  5' p-ATCGCA AGAGAG ACG
                                    CGT TCTCTC TGCCTA-p 5'
```

6.12 Concatenating tags for Sanger and Pyro sequencing

Fri Feb 16 17:18:28 EST 2007

I've had so many problems getting two different adaptors to ligate on opposing ends at the same time, that I've decided to try and do this with Sanger sequencing and/or pyro sequencing. The strategy will be pretty much the same for both. Rather than adding two different adaptors, I'm going to ligate a single adaptor (with the blunt side phosphorylated) to the blunt PET tags (Figure 6.48). I'll probably need to prepare a LOT of PET DNA. Or else add an additional ligation step where I add a blunt 20-30mer and do single-primer PCR to amplify this stuff up. I'd prefer to skip PCR.

For pyrosequencing I'll try to time the ligation reaction so that the concatamers are 300-500bp. For the sanger sequencing, I'll aim for >= 800bp inserts.

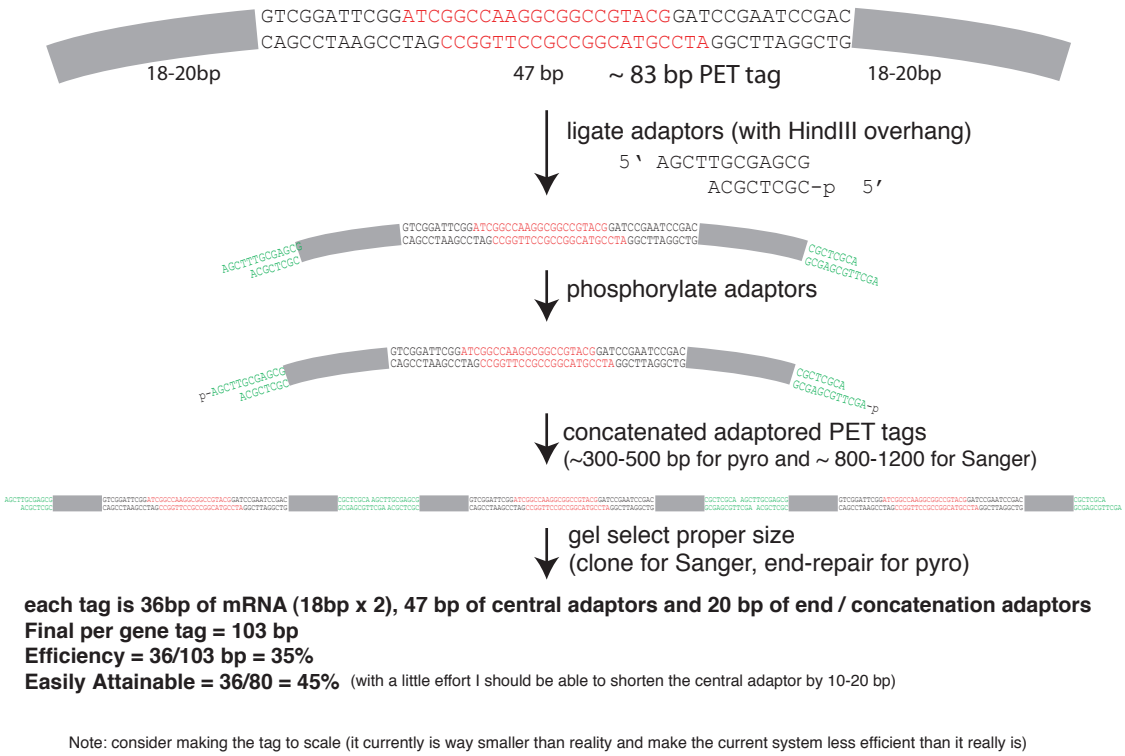


Figure 6.48: PET tags will be adaptored and concatenated so that multiple tags can be read with a single sequencing read (6-10 tags per Sanger, 1-4 per pyro).

6.12.1 The adaptor sequence for concatenation

Fri Feb 16 17:26:33 EST 2007

I'm making a HindIII based adaptor (see below). Only one end is phosphorylated to prevent concatamers. All that should be possible is a blunt/blunt ligation of the adaptors to themselves

plus an adaptor-PET-adaptor ligation (the one we want). The former will be appx 20 bp and the latter should be appx 105 bp. The critical parameters will be the ligation time and the amount of ligase, but they should be easier to optimize than before when I was trying to add two adaptors at the same time. Since the adaptors are so short, I want to do the ligations at 16C. The increase in length from 85 bp to 105 bp should be enough for me to gel select the correct ones. Plus the ones that aren't adaptored will not concatenated nearly as fast in the subsequent concatenation step. As with the adaptoring of the cDNA, I will use a vast excess of adaptor and run the reaction a long time.

```
HindIII concatenation adaptor
5' AGCTTGCGAGCG
       ACGCTCGC-p 5'
```

Brief Update *Fri Mar 23 13:13:54 EDT 2007*: I've designed a new HindIII adaptor. Using this one, I might not need to do the initial end-repair reaction (see new design below).

```
HindIII concatenation degenerate adaptor
5' AGCTTGCGAGCGNN
       ACGCTCGC-p 5'
```

6.12.2 Preparing cDNA for the concatenation approach

Growing cells Mon Feb 19, 2007

I grew cells from a 1:100 dilution of overnite *E. coli* MG1655 culture in LB. Three samples were grown in a baffled flask with 20 ml of LB, and three additional samples were in 20 ml LB + 75 ng/μl norfloxacin antibiotic (Figure 6.49).

I put 2.5 ml of OD 600 0.5 in 5 ml of RNAprotect [Qiagen], vortexed 5 sec, incubated 5 min, and spun down 12 min at 4000 rpm. The pellets were put in the -20C.

Brief Conclusions: No problems so far, as expected the growth of the cells in norfloxacin started to slow after 0.5 OD600 (Figure 6.49).

making ds cDNA from RNAprotect pellet Tue Feb 19, 2007

cDNA was made from samples 1 (LB) and 4 (LB + nor) according to the protocol in section C.9 on page 427.

These are the spec readings for SAMPLE POINT A (total RNA + contaminating genomic DNA: 100 μl):

Sample	DNA (ng/ul)	260/280	260/230	total yield
1	587.0	1.96	2.14	58.7 μg
4	1275.2	2.18	2.04	127.5 μg

These are the spec readings for SAMPLE POINT B (total RNA *without* contaminating genomic DNA: appx 35 μl):

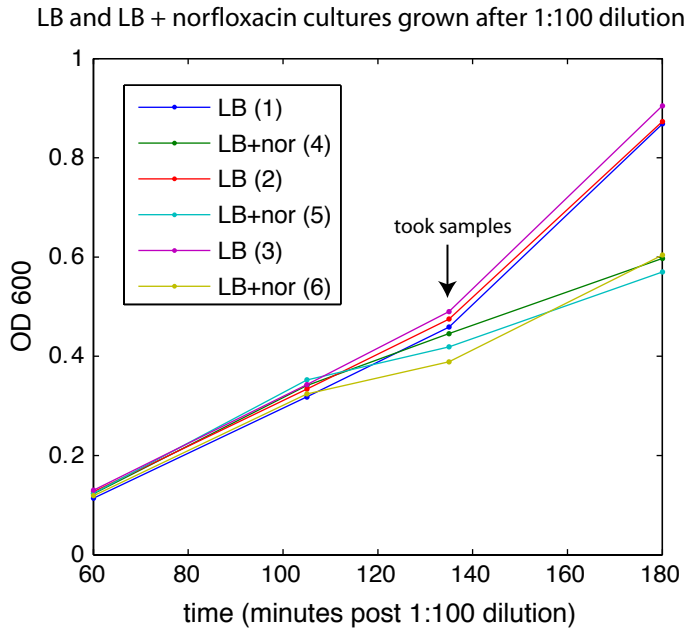


Figure 6.49: Cells were grown after 1:100 dilution from an overnight culture. Samples 1-3 were grown in 20 ml LB, samples

Sample	DNA (ng/ul)	260/280	260/230	total yield
1	1559.1	2.09	2.26	54.6 μg
4	1704.3	2.09	2.26	60.0 μg

Used appx 6 μl (10 μg) of each sample for the MICROBExpress mRNA kit.

These are the spec readings for SAMPLE POINT C (mRNA appx 16 μl):

Sample	DNA (ng/ul)	260/280	260/230	total yield
1	171.6	2.17	2.14	2.7 μg
4	222.2	2.17	2.14	3.6 μg

Brief Conclusions: The RNA samples look pretty good (Figure 6.51). 1a, 4a, 2b, 4b are just what I expect to see. 1c and 4c, post-rRNA removal is still a little disappointing. Once again, the 16S rRNA is removed well, but the 23S is not. I think that this time 23S is a little better removed than last time (Figure 6.12) where the 23S band was even stronger. The only difference this time is that I used water-baths for the MICROBExpress kit rather than heat blocks.

These are the spec readings for SAMPLE POINT D (double stranded, end-repaired cDNA: appx 30 μl).

Sample	DNA (ng/ul)	260/280	260/230	total yield
1	187.8	3.18	3.67	5.6 μg
4	277.5	3.12	3.83	8.3 μg

Brief Conclusions: The 260/280 and 230/260 values were really high here????

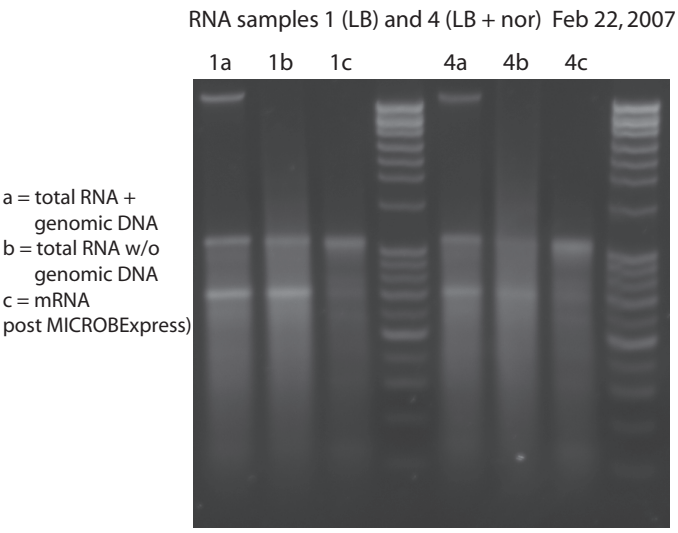


Figure 6.50: RNA was extracted with an RNAeasy kit (a), treated with LiCl and TURBO DNA-free [Ambion] (b), and run through the MICROBExpress kit (c)

size selecting adaptored-cDNA Feb 22, 2007

I ran the adaptoed cDNA on a low-melt agaroase gel (Figure 6.51). And cut the bands for 500-1500 and > 1500 for both cDNA sample 1 and cDNA sample 4.

agarase preparation of cDNA Feb 23, 2007

I digested the 4 gel slices with β-agarase [NEB] using the NEB protocol. I think there was too much agarose, because it really didn't seem to digest too well. There was a lot of gel in the bottom after spinning down. I should call NEB and see if this is normal. I cleaned the agarases digestions with isopropanol. Then I eluted into 15 μl TE.

circularize size-selected cDNA Feb 23, 2007

I used 5 μl of the cDNA (1/3), 120 ng of the circularizer DNA (pre-annealed in STE), and 1 μl ligation buffer in a 10 μl ligation at 16C for 2hr. I did 2 reactions with cDNA sample 1 big (i.e. >1500bp), I used circularizer 2 (which creates a 77mer). With cDNA sample 4 big, I used circularizer 1 (the original, which creates an 83mer).

RCA the circularized cDNA Sat Mar 24, 2007

5.25 μl 10x buffer, 2 μl dNTP, 5 μl template DNA, 35.25 H_2O , 2.5 μl hexamer. heat to 95C for 5 min, cool to 4C, transfer to ice. add 2.5 μl ϕ29 polymerase, incubate at 30C for 12 hr, heat deactivate 10 min at 65C.

Mon Mar 26, 2007

I EtOH precipitated the RCA and resuspended the DNA into 200 μl of TE. The yields were (number each cDNA sample, b = big = >1500bp):

Sample	DNA (ng/ul)	260/280	260/230	total yield
RCA 1b	359.8	1.78	2.47	72 μg
RCA 4b	340.1	1.80	2.48	68 μg

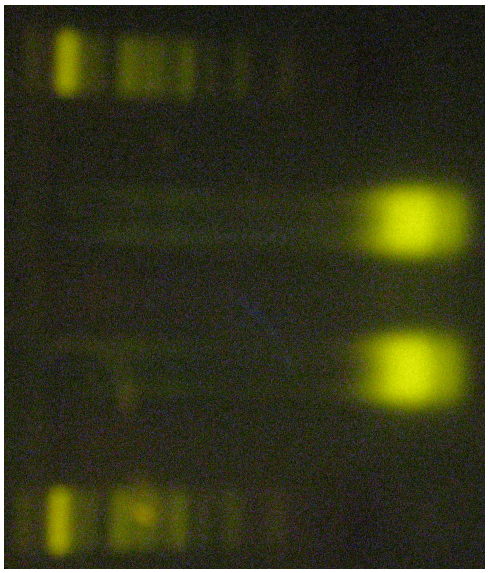


Figure 6.51: Thur Mar 22, 2007. cDNA samples 1 and 4 on a 1% SeqPlaque lowMT gel and two lanes of 1kb ladder. I know, pretty much all you can see are the adaptors. Maybe next time I'll use the Qiagen PCR kit to help remove them before running the gel.

Brief Conclusions: Mon Mar 26, 2007 The yields above are about half of what I expected based on the previous result (see section 6.6.2 on page 270). However, I realize now that I forgot to do the exonuclease reaction to remove the non-circular DNA. 60 μl of each of the above RCA = 20 μg , which is what I want for my MmeI digestions.

MmeI digestion of the RCA'd circularized cDNA Mon Mar 26, 2007

I ran the following digestion: 60 μl (20 μg) RCA, 10 μl MmeI, 4 μl SAM (diluted 1/20), 10 μl NEB4, 16 μl H_2O at 37C for 15 min and 10 min at 65C to deactivate the MmeI. I EtOH precipated the rxn and eluted into 15 μl TE.

Nusieve of MmeI digested RCA Tues Mar 27, 2007

I ran the MmeI digested 20 μg of RCA'd cDNA for samples 1b and 4b on to a 3.5% TBE gel (Figure 6.52).

Brief Conclusions: Although it's hard to see in Figure 6.52, the two barcodes were faintly visible and the correct sizes (i.e. the new one looked appx 5bp shorter).

RCA the circularized cDNA Tues Mar 27, 2007

In case it mattered, I added the Exonuclease I and the Exonuclease III to the circularized cDNA to remove non-circular DNA. The protocol was: add 2 μl Exonuclease I and 0.4 μl ExonucleaseIII, incubated 45 min at 37C, heat-deactivate 80C for 20 minutes.

After removing the linear DNA, I make the following RCA rxn: 5.25 μl 10x RCA buffer, 4 μl dNTP, 5 μl template, 33.25 H_2O , 2.5 μl hexamer; incubated 95C for 5 min, to ice, added 2.5 μl ϕ29, 30C for 12 hr, 10 min 65C to deactivate.

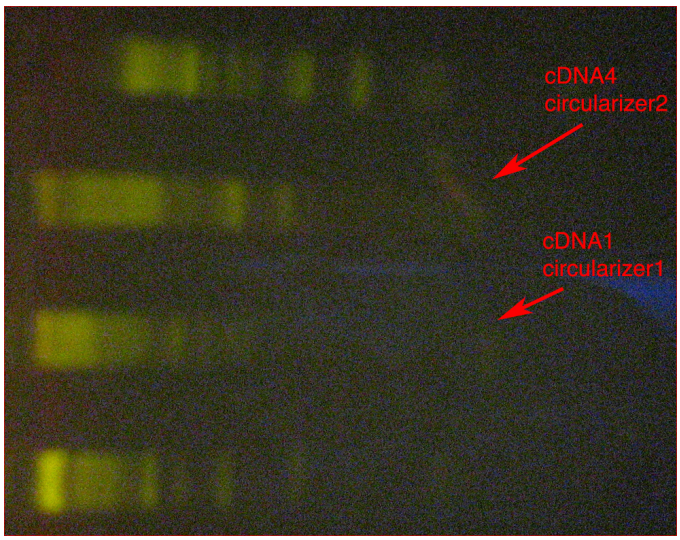


Figure 6.52: 3.5% Nusieve gel with sybr safe run at 80V for 90 minutes. Gel shows the MmeI digested RCA products from two cDNA samples.

Wed Mar 28, 2007

After an EtOH, spec'd the two RCA reactions to see the following yields:

Sample	DNA (ng/ul)	260/280	260/230	total yield
RCA 1b	2722.2	1.85	2.35	544 μg
RCA 4b	2670.4	1.85	2.35	534 μg

Brief Conclusions: Wow, much better yield than the last RCA. Almost 10x higher yield. I changed two parameters from last time: 1) I doubled the amount of dNTP, 2) I removed the linear DNA. You wouldn't think that doubling dNTP would give an order of magnitude more yield, so I guess digesting the linear DNA away is really important.

6.12.3 Ligating tags into circles for RCA amplification

I want to be able to ligate the PETs into a circle that is >= 500 bp. The circle provides an easy way to amplify concatenated tags, since I think it is unlikely to be able to meet the 454 pyrosequencing requirement of 5 μg of *clean* DNA. So the protocol will go: 1) make tags, 2) concatenate into big circles, 3) amplify into long, linear RCA DNA, 4) send to 454 for shearing and sequencing. The downside of this requirement is that you end up with partial PET sequences for the start of your sequencing read (depending on where the shearing breaks your DNA).

The persistence length of DNA is around 150 bp under normal salt conditions. This means the circles are unlikely to be less than 150 bp in length, but we want to push that towards larger circles. The basic approach comes from some suggestions from Ravi Sachidanandam (my old boss at CSHL and a former physicist that gets a kick outta this kinda thing).

Ravi's rules for biasing towards larger circles The overall strategy of make larger circles it to create conditions where each PET end finds another PET end more often than it finds its own end (self-end).

- higher concentration of PETs (probably done by decreasing the ligation rxn volume) (if more PETs are in a smaller space, their ends are more likely to meet each other)
  - start with small ligation volume (e.g. 10ul)
  - increase volume a LOT when you are ready to coerce to circles (e.g. 100-1000ul with appropriate buffer)
- lower temperature (DNA moves around less and is therefore less likely for the self-ends to meet)
  - going to much below 16C will likely make the ligation to efficient
  - maybe ligate at 16C (or perhaps even 12C) then raise the temp to 25C when you want to coerce them into circles
- lower NaCl (below 0.1M) (too low salt will also inhibit ligation) (low salt makes the DNA stiff, so that the self-ends will be unlikely to meet)
  - already tried 1/2 and 1/4 ligase buffer and the ligation seems to work fine, I don't know if it effects the circle size yet though
  - use 1/4 ligation buffer, then increase buffer to 1x to coerce to circles
- increase viscosity (more viscous solutions also act to keep the DNA from bending around too much)
  - checkout NEB recommendations for PEG additions to ligation
  - dilute to lessen the PEG effect

tests on fake tags (coming from amplified pUC19 with HindIII overhangs)

I'm using a 77mer and an 83mer (the current sizes of the tags I'm using) to try and approximate the real results of ligating the tags without having to go through the long process of making and gel-selecting the tags each time. I'm using 2 tags of each size amplified from regions of pUC19 that 1) do not have a HindIII site (since HindIII is the adaptor sequence) and 2) are not in regions that might have strong secondary structure (e.g. promoters and transcription terminators). The features of pUC19 are below:

```
Features:
  469- 146      lacZ alpha CDS (start 469, complementary strand)
  519- 514      Plac promoter -10 sequence (TATGTT)
  543- 538      Plac promoter -35 sequence (TTTACA)
  575- 563      CAP protein binding site
  396- 452      multiple cloning site (EcoRI-HindIII)
 1455- 867      origin of replication (counterclockwise)
                  (RNAII -35 to RNA/DNA switch point):
 1273-1278        RNAI transcript promoter -35 sequence (TTGAAG)
 1295-1300        RNAI transcript promoter -10 sequence (GCTACA)
 1309-1416        RNAI transcript
 1419- 867        RNAII transcript (complementary strand)
 1434-1429        RNAII transcript promoter -10 sequence (CGTAAT)
```

```
 1455-1450        RNAII transcript promoter -35 sequence (TTGAGA)
 2486-1626      beta-lactamase (bla; amp-r) CDS
                  (start 2486, complementary strand)
 2486-2418      beta-lactamase signal peptide CDS
                  (start 2486, complementary strand)
      2521      bla RNA transcript start (complementary strand)
 2535-2530      bla promoter -10 sequence (GAGACA)
 2556-2551      bla promoter -35 sequence (TTCAAA)

 1 LEFT PRIMER       2046   20   60.21   60.00  4.00  0.00 GGTTAGCTCCTTCGGTCCTC
   RIGHT PRIMER      2128   20   59.72   50.00  8.00  1.00 TATGCAGTGCTGCCATAACC
   PRODUCT SIZE: 83, PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 1.00

 2 LEFT PRIMER       1878   20   59.97   50.00  6.00  1.00 TTGCCGGGAAGCTAGAGTAA
   RIGHT PRIMER      1960   20   60.60   55.00  5.00  2.00 GTGACACCACGATGCCTGTA
   PRODUCT SIZE: 83, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 2.00

 3 LEFT PRIMER       1878   20   59.97   50.00  6.00  1.00 TTGCCGGGAAGCTAGAGTAA
   RIGHT PRIMER      1959   20   60.74   55.00  3.00  2.00 TGACACCACGATGCCTGTAG
   PRODUCT SIZE: 82, PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 3.00

 4 LEFT PRIMER       1216   20   60.14   55.00  6.00  2.00 GCAGCCACTGGTAACAGGAT
   RIGHT PRIMER      1298   20   60.65   55.00  4.00  0.00 TAGCCGTAGTTAGGCCACCA
   PRODUCT SIZE: 83, PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 1.00

ORDERED 1 and 2 (I refer to these as 83.1 and 83.2 in my experiments)

AND a 77 bp product almost exactly like 1
   LEFT PRIMER       2048   18   55.43   55.56  4.00  0.00 TTAGCTCCTTCGGTCCTC
   RIGHT PRIMER      2124   20   60.69   50.00  4.00  2.00 CAGTGCTGCCATAACCATGA
   PRODUCT SIZE: 77, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 2.00

AND a 77 bp product almost exactly like 2
   LEFT PRIMER       1879   20   59.61   55.00  4.00  0.00 TGCCGGGAAGCTAGAGTAAG
   RIGHT PRIMER      1955   19   60.28   52.63  3.00  2.00 ACCACGATGCCTGTAGCAA
   PRODUCT SIZE: 77, PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 1.00

ORDERED 1 and 2 (I refer to these as 77.1 and 77.2 in my experiments)

add AATTAAGCTT to LEFT
add ATATAAGCTT to RIGHT

Final sequences (minus the HindIII adaptor) are:

77.1
TTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTG

77.2
```

TGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGT

83.1
GGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATA

83.2
TTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCAC

The "add" sequences above were added to the 5' end of the left and right primers respectively. They allow the addition of a HindIII site to the 83mer and the 77mer just as we'll have with the real PETs. In addition they add 4 bp just to allow the HindIII to cut. The cutting enzymes will leave the phosphates, so that the cut PCR products can just be ligated directly after cutting (perhaps with an initial clean up).

PCR amplifying the pUC19 based tags Thur Mar 8, 2007

I made 200 μl rxns of each of the four primer pair sets. I cleaned them with Qiagen PCR purification kits and eluted into 30 μl . The yields were:

Sample	DNA (ng/ul)	260/280	260/230	total yield
77.1	73.8	1.92	2.25	2.2 μg
77.2	76.4	1.89	2.11	2.3 μg
83.1	78.5	2.01	2.05	2.4 μg
83.2	72.7	1.65	2.50	2.2 μg

Brief Conclusions: Pretty low yields considering I used 200 μl for my PCRs (no in one tube, that's 2 x 100 μl rxns. I don't know if the column is not catching the short DNA well or if the short DNA PCR just doesn't result in as much product (e.g. if PCR give you a certain number of molecules than a 200mer product should yield 1/10th the amount of DNA [by weight] of a 2000bp piece).

83.2 tag digestion try 1 Sat Mar 10, 2007

I digested all 30 μl of the PCR product for tag 83.2 using the following rxn conditions: 30 μl cleaned PCR, 4 μl NEBuffer2, 0.5 μl HindIII, 5.5 μl H_2O . 37C for 30 min, followed by heat deactivation. The rxn was cleaned with a Qiagen PCR purification kit with the following yield (into 34 μl of EB):

Sample	DNA (ng/ul)	260/280	260/230	total yield
83.2	36.8	1.98	2.30	1.3 μg

83.2 tag ligation try 1 Sat Mar 10, 2007

I ran three ligation reactions just to try things out the rxns were all 20 μl total volume with 10 μl of DNA (368 ng) and 1 μl of T4 ligase. All rxns were at 16C with a final heat activation at 65C for 10 minutes.

The differences for each were:

1. rxn 1: ligation for 15 minutes with 100% T4 buffer
2. rxn 2: ligation for 60 minutes with 100% T4 buffer
3. rxn 3: ligation for 60 minutes with 1/4 diluted T4 buffer

Wed Mar 14, 2007

I ran the three ligations on a 2% TAE gel (Figure 6.53).

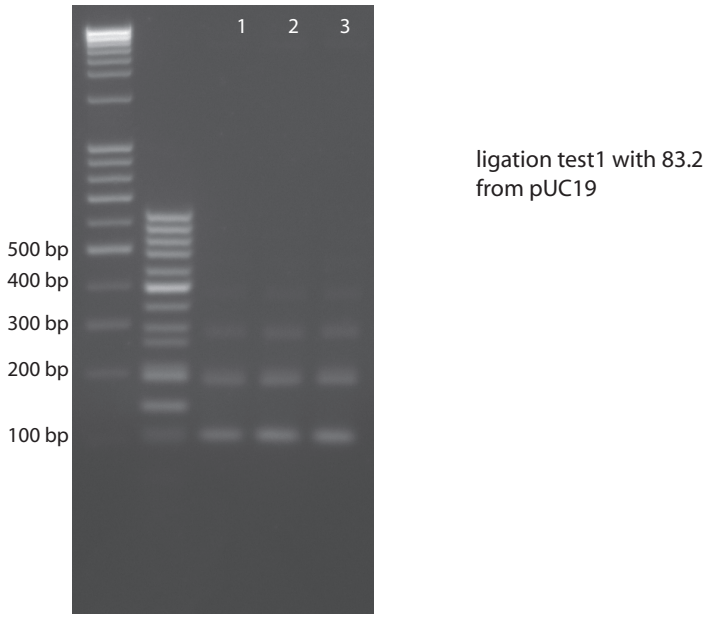


Figure 6.53: 83.2 was amplified from pUC19, cut with HindIII, and ligated in 3 different conditions. The resulting concatentations are multiples of the original length of around 93 bp (83 bp tag + 10 bp for the HindIII site on each end).

Brief Conclusions: First concatenation doesn't look too bad (see Figure 6.53). Looks like the DNA is all linear (just a guess, but I figure if it were non-linear the band lengths wouldn't be just multiples of the original 93 mer, since circular DNA migrates differently). The only worrying thing is that the 83.2 unligated band looks a little bit too long (it appears longer than the 100 bp band in the ladder, but it should be 7bp or so *shorter*). Also, in the next attempt I definitely need to ligate longer, and hopefully I'll have enough DNA to try the Exonuclease step as well.

PCR amplifying the pUC19 based tags try 2 Sat Mar 10, 2007

I ran two 500 μl PCRs: (0.5 μl pUC19 = 250 ng), 10 μl primer 250 μl Taq master mix, 239 μl H_2O . The reactions were concentrated with EtOH and resuspended into 15 μl TE. The primer pair used was 83.2

cutting/ligating the pUC19 based tags try 2 Mon Mar 26 17:30:08 EDT 2007

I digested the first of the two 500 μl PCR reactions: 15 μl of tag 83.2, 2 μl buffer, 0.5 μl HindIII, and 2.5 μl H_2O . Incubated at 37C for 30 min and deactivated the HindIII for 20 min at 65C.

Sample	DNA (ng/ul)	260/280	260/230	total yield
9.1	36.8	1.81	2.10	0.28 μg

Brief Conclusions: I think the Qiagen columns, poor as they are, might be working better than EtOH with this short mer (maybe use Glyco-blue [Ambion] or switch back to Qiagen).

PCR amplifying the pUC19 based tags try 3 Tue Mar 27, 2007

I ran a 600 μl PCR (300 μl Master mix, 12 μl pUC19 (12 ng), 15 μl primer (250 nM), 273 μl H_2O . Ran 35 cycles (first 5 annealed at 54C, last 30 cycles annealed at 59C); I cleaned the entire reaction and ran it through a Qiagen PCR purification kit (I used the vacuum and just kept adding more PBI buffer mixed with PCR product until I was out of PCR product). The yield was:

Sample	DNA (ng/ul)	260/280	260/230	total yield
83.1	107.3	1.89	2.35	3.3 μg

I cut all 3 μg of the PCR product in a 40 μl digestion with 0.5 μl of HindIII for 30 min at 37C and 20 min at 65C. I cleaned up the digestion with a Qiagen PCR clean up kit to get the following yield:

Sample	DNA (ng/ul)	260/280	260/230	total yield
83.1 (post cut)	68.8	2.04	2.27	2.1 μg

Brief Conclusions: Compared to try1 where I also used the Qiagen cleanup, the yields were higher. But in try1, I used only 200 μl for the PCR! Maybe I wrote that down wrong? Tomorrow, I'm going to try 400 μl and 1 ml PCR. I could be that somehow these short DNA fragments are exhausting the Qiagen column? In which case, I'll have to pick the optimal PCR volumn and just run a bunch of them.

ligating the 83.1 try 3 Tue Mar 27 19:40:00 EDT 2007

I ran 3 ligations (similar to try 1, but with more DNA, longer incubations, and more T4 ligase):

1. rxn 1: ligation for 2 hours with 100% T4 buffer and 400 U T4 ligase
2. rxn 2: ligation for 12 hours with 100% T4 buffer and 400 U T4 ligase
3. rxn 3: ligation for 12 hours with 100% T4 buffer and 800 U T4 ligase

Wed Mar 28, 2007

All 20 μl of the ligations were run on a 2% TAE gel for 50 minutes (Figure 6.54).

Brief Conclusions: The 2hr incubation time worked much better than the 1hr and the 15 minute incubation times (compare previous result in Figure 6.53 with the current result in Figure 6.54) The strongest band is still the PCR tag. I'm surprised the 2x ligase didn't help remove more of that band. Next time I want to try the high concentrate ligase that we have. Also, it is possible (unlikely?) that the things are circularizing at that size already? Maybe that's why I can't get rid of that band? I don't know. I'll try high conc ligase AND exonuclease (but maybe not together) on the next round.

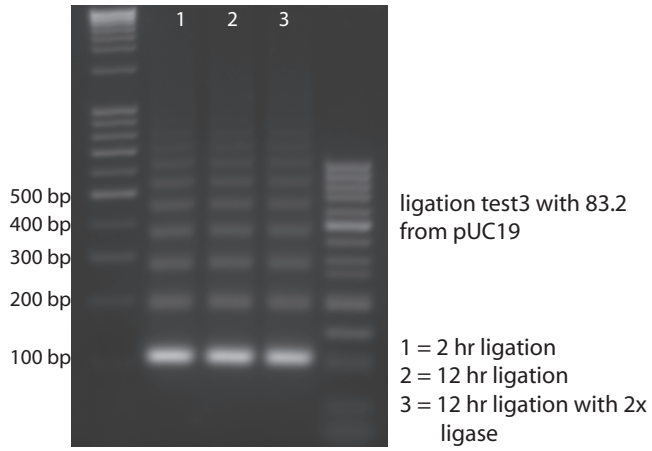


Figure 6.54: 83.1 was amplified from pUC19, cut with HindIII, and ligated in 3 different conditions. The resulting concatentations are multiples of the original length of around 93 bp (83 bp tag + 10 bp for the HindIII site on each end).

PCR amplifying the pUC19 based tags try 4 Wed Mar 28 18:19:32 EDT 2007

I ran 2x 1000 μl (1-2) and 2x 500 μl (3-4) PCR reactions just as in try 3 to try and figure out where I'm losing DNA.

Sample	DNA (ng/ul)	260/280	260/230	total yield
83.1 (1)	40.7	1.86	1.63	1.2 μg
83.1 (2)	60.6	1.82	1.82	1.8 μg
83.1 (3)	24.3	2.36	1.50	0.73 μg
83.1 (4)	25.4	1.85	1.40	0.76 μg

Brief Conclusions: This sucks. Come on, how hard can it be to purify this little piece. I might try using EtOH again, but with Glycoblue. Alternatively, I might make a 160 mer with a HindIII site in the middle, then I'd make a double tag and cut it (3x: once on each end and once in the middle).

cutting and ligating PCR try 4 Fri Mar 30 15:22:17 EDT 2007

Based off the results in Figure 6.59, we need to work harder to coerce these ligations into circles. In this experiment, I'm going to continue with the 2 hr ligation. After those 2 hrs, I'm going to dilute the ligation 1:10 (to make my 40 μl rxn 400 μl). This lower concentration should help push things towards circles rather than ligating with each other. I'm also going to add more ligase at the same time in case the initial ligase is exhausted. Last, I'm going to try this using normal and high concentration T4 ligase (400U and 2000U).

I digested the two 1ml PCR reactions from try 4 (81.1 (1) and 81.1 (2)) in separate 40 μl reactions with 0.5 μl of HindIII at 37C for 45 minutes followed by at 20 min heat deactivation at 65C. The two digestions were cleaned with a Qiagen PCR clean up kit and eluted into 30 μl of EB buffer. 4 μl of T4 ligase buffer was added to both elutions for a 40 μl ligation. 81.1 (1) used 400 U of T4 ligase (1 μl of normal concentration) and 81.1 (2) used 2000 U of T4 ligase (1 μl of high concentration). The reactions were run at 16C for 2 hrs and the ligase was NOT deactivated.

Then I diluted the ligations 1:10 in H_2O + T4 ligase buffer (this created one 400 μl reaction for each ligase concentration and required breaking the reactions into two PCR tubes. To each 400 μl ligation reaction one additional microliter of ligase was added (normal conc for 81.1 (1) and high conc for 81.1 (2)). The ligations were run for 2 hr at *25*C (hopefully the higher temperature will aid circularization). The ligase was then heat deactivate at 65C for 10 minutes.

Sat Mar 31 16:48:12 EDT 2007

The two 400 μl ligation reactions were concentrated with EtOH precipitation and 1 μl of glycoblue. The tubes were placed at -20C for 20 minutes and then at -85C for 20 minutes. They were spun for 20 minutes at 4C followed by washing with 750 μl 70% EtOH. The pellet was dried for 5 minutes and resuspended in 30 μl of TE.

To one-half of the two ligation reactions (and to 250 ng of a pUC19 control), I added 2 μl Exonuclease I and 0.4 μl Exonuclease III. I added 1.8 μl of NEBuffer1. The 3 rxns were digested for 30 min at 37C followed by 20 min at 80C. All 6 reactions (including the two pUC19 controls) were run on a 2% gel 6.55.

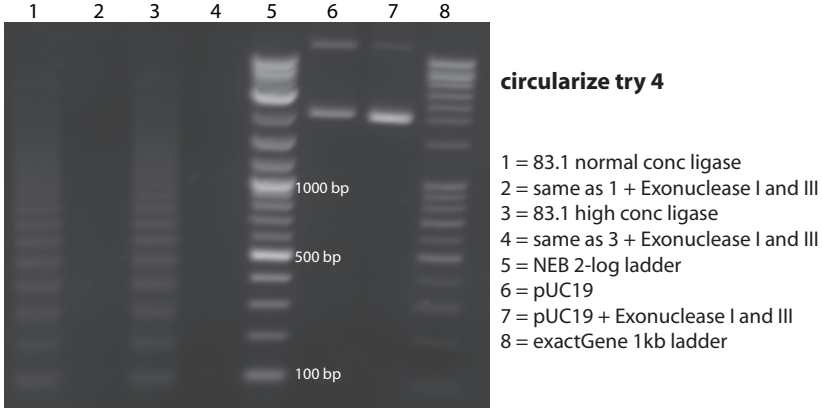


Figure 6.55: 2% gel run for 50 minutes.

Brief Conclusions: I think the problem is actually the opposite of what I thought it was. We can't get the damn DNA to circularize. All it will do is concatenate into longer linear fragments (see attached image). So I guess now I just try the opposite of all the things we were doing to push towards linear: 1) lower the concentration of DNA (that's what I tried in the attached picture. I lowered it 1:10, apparently that's not enough. Unfortunately, 1:18 is about as low as I can go. 2) increase temperature this one I also max'ed out on the attached gel (I increased T by 9C); I might be able to switch to the Taq ligase (like the Taq polymerase it works at really high T), but won't this also increase diffusion and make the linearization more likely? 3) increase salt, this might be possible, I'll have to check 4) decrease viscosity, not really possible, I never added anything viscous

PCR amplifying the pUC19 based tags try 5 Thu Mar 29, 2007

Ran 2 x 500 μl PCRs and 2 x 200 μl PCRs (using same conc of puc19 and primer as before). I cleaned them up with ethanol precipitation using glycoblue as a carrier. Hopefully, this carrier increases my yield (which with my last EtOH of this short piece was appx 0 ng/μl). With one of the duplicates I used 1 μl of carrier and with the other I used 5 μl . I let the reaction stay at -20C

for 30 min and at -85C for 30 min. I spun at 4C for 30 min, washed with 750 μl of 70% EtOH, spun 10 more min at 4C. Dryied the pellet for 5 minutes and resuspended in 30 μl of TE. I also made two blank of 30 μl TE with 1 μl and 5 μl of glycoblue respectively. The yields where:

Sample	DNA (ng/ul)	260/280	260/230	gel estimate
blank (1 μl glyco)	6.6	1.76	0.90	0
blank (5 μl glyco)	41.3	1.73	0.93	0
83.1 (500 μl ; 1 μl glyco)	432.2	1.60	2.10	56.4
83.1 (500 μl ; 5 μl glyco)	595.0	1.71	1.92	66.8
83.1 (200 μl ; 1 μl glyco)	386.9	1.63	2.19	24.5
83.1 (200 μl ; 5 μl glyco)	539.4	1.66	1.96	30.6

The values looked fishy, so I ran 1 μl of each reaction on a gel (I ran the blanks too, but as expected they were not visible on the gel) for gel based quantification. The software estimated values for this quantification are shown last column in the table above.

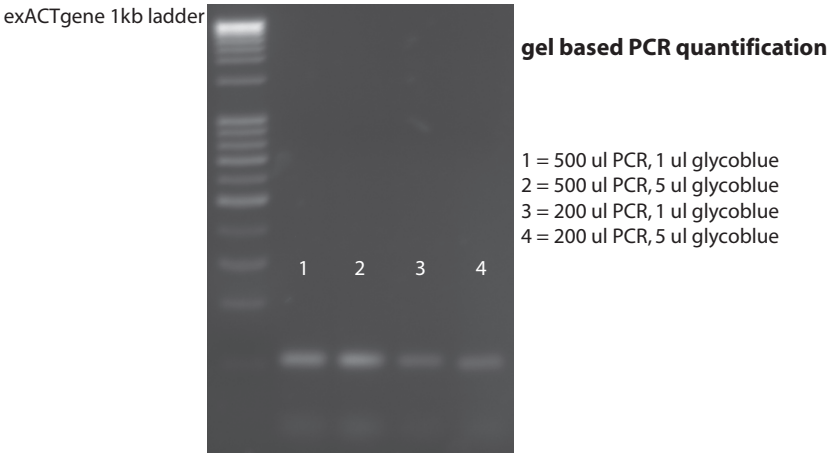


Figure 6.56: 1 μl of each EtOH/glycoblue precipitated PCR rxn was run on a gel for quantification (since the spec seemed inaccurate, perhaps due to dNTPs and primers).

Brief Conclusions: The gel-based quantification was definitely better than the spec. However, I think the gel-based values are pretty large underestimates, because the bands get fuzzy at the end. I'm going to switch to the NEB 2-log ladder, which provides bands with a larger dynamic range that should help quantification.

cutting and ligating the pUC19 based tags try 5 Apr 4, 2007

I cut 83.1 (500 μl ; 1 μl glyco) with HindIII. I cleaned up the digestion with a Qiagen PCR cleanup kit and got:

Sample	DNA (ng/ul)	260/280	260/230	yield μg
cut 83.1 (500 μl ; 1 μl glyco)	46.7	1.85	2.13	14.0

I'm divided the digested tags into 3 groups each using a different two stage ligatino procedure

1. (stage 1) 10 μl cut DNA, 2 μl T4 buffer 8 μl H_2O , 1 μl T4 16C for 2 hr; (stage 2) add 179 μl H_2O + T4 buffer, 1 μl T4 ligase 16C for 2 hr

2. (stage 1) 10 μl cut DNA, 2 μl T4 buffer 8 μl H_2O , 1 μl T4 16C for 2 hr; (stage 2) add 179 μl H_2O + Taq ligase buffer, 1 μl Taq ligase 45C for 2 hr

3. (stage 1) 10 μl cut DNA, 2 μl Taq ligase buffer 8 μl H_2O , 1 μl Taq ligase 45C for 2 hr; (stage 2) add 179 μl H_2O + Taq ligase buffer, 1 μl Taq ligase 45C for 2 hr

I digested half of the each of the 3 reactions with exonuclease I and III. I ran them all on a 2% agarose gel (Figure 6.57). I cut out the bands at around 500 bp from the non-exonuclease gel.

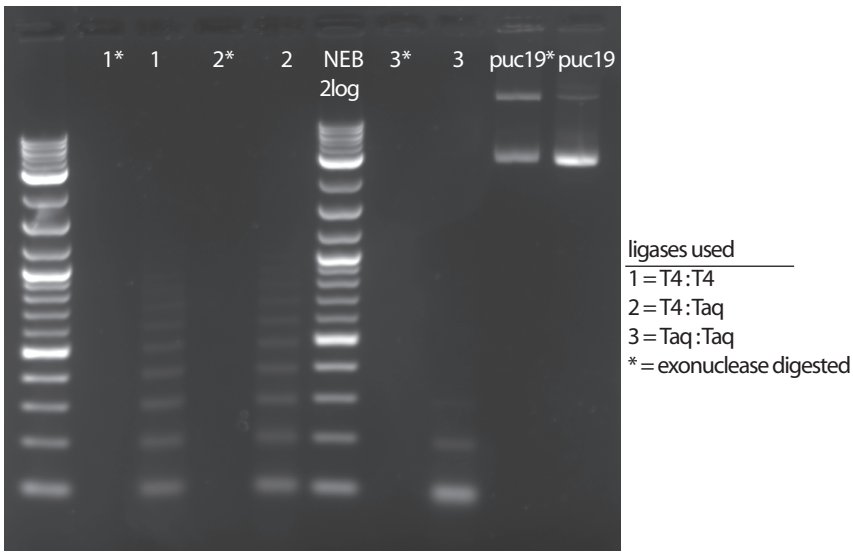


Figure 6.57: 2% gel run for 45 minutes.

Apr 5, 2007

I cleaned two gel slices (lanes 1 and 2 in Figure 6.57) using a Qiagen column-based gel extraction kit and a QiaexII gel extraction kit (to see which method was better).

I then made the assumption that if I added ligase, at least a few of the cut pieces would ligate into a circle. Allowing me to amplify my tags even if I wasn't able to see the circles on a gel (because so far they are way to few relative to the uncircularized pieces). I used the entire gel clean up reactions in a 40 μl ligation with 4 μl T4 buffer, 1 μl T4 ligase, and 5 μl H_2O (I did one of these reactions for each of the two samples). I incubated 12hr at 16C and heat deactivated at 65C for 10 minutes.

I took the (hopefully) circularized ligation and digested the linear fragments using Exo I and Exo III. I used 5 μl of this in an RCA reaction to amplify the circles.

Sat Apr 7, 2007

I did an EtOH precipitation of the RCA.

Sun Apr 8, 2008

I spec'd the RCAs:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**yield (μg)**
83.1 Qiaex	2388.9	1.81	2.29	119.4
83.1 Qiagen column	2694.2	1.77	2.23	134.7

I sheared 4 μl of each RCA (appx 10 μg) in 125 μl TE at 10% power. I cleaned up the reaction with a Qiagen PCR cleanup:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**yield (μg)**
83.1 Qiaex sheared	71.4	1.88	2.31	2.42
83.1 Qiagen column sheared	20.3	2.18	2.22	609

I ligated the sheared DNA with 50 ng of blunt (SmaI) cut, dephosphorylated puc19, using 1 μl of concentrated T4 for 10 min at RT, and then heat deactiviated at 65C for 10 min. I cloned this into DH5α competent cells.

Mon Dec 9, 2007

To make sure the RCA and ligation worked (well I know the RCA amplified something because I got a ton of DNA, I just don't know if it amplified the correct thing), I digested 10 μg of the RCA with HindIII. 4 μl RCA DNA, 2 μl NEB2, 13 μl H_2O , 1 μl HindIII. I ran the digestion for 45 min at 37C followed by deactivation at 65C for 10 minutes. I ran all of the digestion reaction onto a 2% agarose gel.

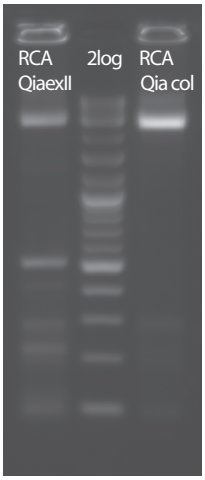


Figure 6.58: 2% gel run for 45 minutes. Unfortunately, I don't see the expected band at 83 bp.

Brief Conclusions: I doesn't look like this worked too well. For sure there was circular DNA in the RCA reaction because I get a huge amount of amplification (perhaps some sorta contaminating circular DNA?). However, when I gut that amplified DNA with HindIII, I should be back to my 83mer. And that doesn't seem to be the case (Figure 6.58). Esther had better luck when she tried the same thing using a one stage ligation and no gel extraction (but there we still don't know if the circles were big enough. perhaps we can size select the circles using the ChargeSwitch PCR cleanup kit?). I'm also considering using a PCR-plug like is used in LM-PCR. I can just something like:

```
5' p-AGCTGCTAGC
         CGATCG

www.flychip.org.uk/protocols/chip/lm_pcr.pdf
http://nar.oxfordjournals.org/cgi/content/full/27/18/e23
```

In excess, the plug would halt the ligation/concatenation (by blunting the ends with an unphosphorylated primer). And it could be used together with the HindIII site to amplify the tags via PCR rather than rolling circle.

`BEGIN WORK BY Esther Rheinbay`
Mar 29, 2007

Exonuclease digestion of linear and circular DNA: do we have ligated circles? Esther just did a 50 μl PCR to get started on this project of trying to create definable sized circular DNA from these 83mers and 77mers. As an initial check to see if the smaller DNA was forming circles during the ligation (and to make sure that the exonuclease really does remove the non-circular DNA), she digested known circular DNA (pUC19) and the ligated 83.2 fragments. She ran the resulting four lanes (cut/uncut for both) on a 2% gel for 45 min (Figure 6.59).

She used 2 μl Exonuclease I and 0.4 μl Exonuclease III, incubated at 37C for 45 min and heat deactivated at 80C for 20 minutes. 100 ng of puc19 DNA was used in the control.

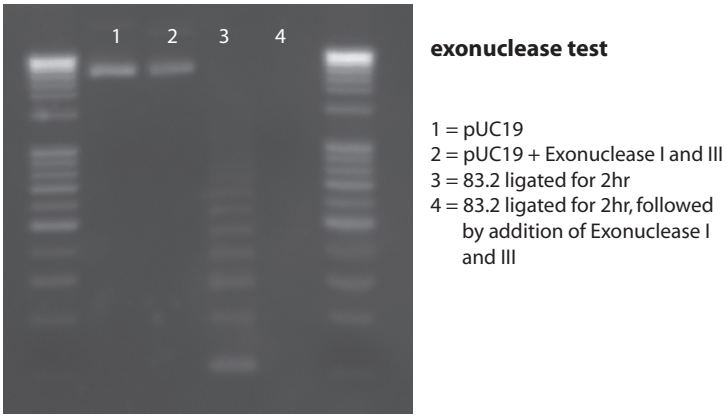


Figure 6.59: 2% gel run for 45 minutes. The perfect example of circular DNA (lanes 1 and 2) surviving Exonuclease I and III digestion (lane 2) while linear DNA (lanes 3 and 4) gets completely digested (lane 4).

Brief Conclusions: A few nice things can be concluded from this gel. 1) the exonuclease trick definitely works as the circular plasmid was just slightly digested (presumably due to nicks and other DNA breaks) while the linear DNA was completely removed. 2) we certainly can get linear DNA higher than the 150bp persistance length of DNA. The problem might even be how to we get the ligated pieces to form circles at all. At least initially, we can try the opposite of the tricks to bias away from circles. So after the initial 2 hr ligation, we can dilute a lot (1:10? or maybe 1:100?), but keep the salt conc high and incubate at higher temp (25C). We'll also need to add more ligase, because it seems to be exhausted after that initial 2hrs.

Esther also ran the following (similar to my try 5 but worked better): digest 83.2 with HindIII, ligate 2 hr with T4 ligase. Digest with Exonuclease I and III to remove all non-circular DNA (she actually skipped this step?). RCA. Digest RCA with HindIII. She did this for two reactions, one

worked and the other failed (Figure 6.60). The unfortunate thing is that she forgot to do the Exonuclease digestion. Would be much more confident in the result if she'd done that.

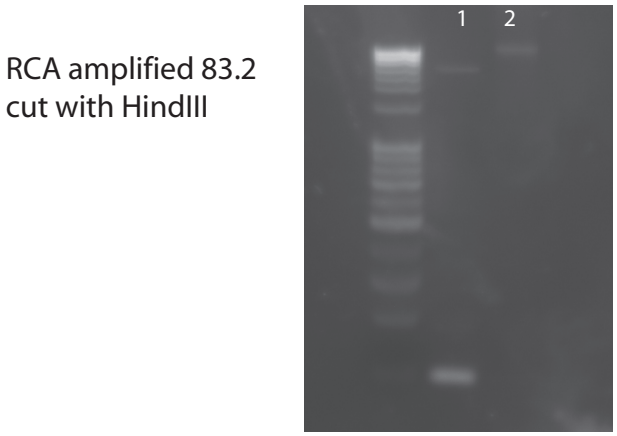


Figure 6.60: 2% gel run for 45 minutes. Seems lane 1 worked and lane 2 didn't.

```
END WORK BY Esther Rheinbay
```

6.12.4 amplification of tags with LM-PCR

Wed May 2, 2007

Previously I've been ligating the tags into a circle for amplification (after which they could be sheared and sequenced). I wanted to try Ligation Mediated PCR (LM-PCR) as an alternative. The plan is to do the initial ligation for 2hr, then to add a blunt-dephosphorylated plug

```
5' p-AGCTGCTAGC
         CGATCG
```

As I mentioned a few sections before, in excess, the plug would halt the ligation/concatenation (by blunting the ends with an unphosphorylated primer). And it could be used together with the HindIII site to amplify the tags via PCR rather than rolling circle.

Wed May 2, 2007

I amplified 77.1, 77.2, 83.1, and 83.2. I also ran a separate reaction with the 83.1 only, so I could look for discrete bands whereas mixing the two sizes might make things a little smeary. I did a 2 hr ligation with 1 μl of T4 at 16C. I then added 500 ng (1 μl in STE buffer) of annealed plug plus 1 μl of T4 buffer and 1 μl of T4 ligase and I incubated another 2 hr at 16C followed by deactivation at 65C.

Thur May 3, 2007

I ran a PCR using the single primer (1 μl of 10 μM). I used 5 μl ligation product and annealed at 60C. The PCR was for 30 cycles. Unfortunately, yesterday I forgot to make a control with no plugs. I ran the ligation and the LM-PCR from the ligation onto an agarose gel (Figure 6.61).

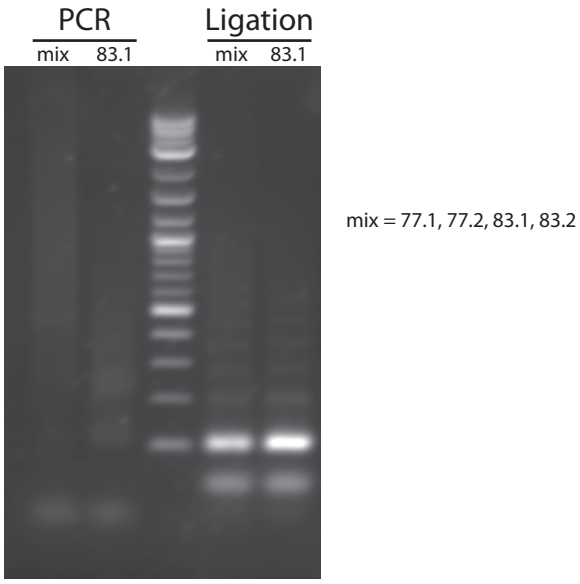


Figure 6.61: 1.5% gel run for 45 minutes.

Brief Conclusions: Looks way to smeary (Figure 6.61). I don't think that was succesful. The circular method seems much better for the moment.

6.13 Further removal of rRNA

Tues Jul 3, 2007

I want to have fewer rRNA samples in my sequencing results. Without using MicrobeExpress, the sequences were virtually 100% rRNA. After one round of MicrobExpress I sequenced a couple of real genes for the first time. Still 80% of the reads were 23S (though 16S seemed to be be largely removed on the gel and from the sequencing results; see section 6.5 on page 267 for more details). I want to try and remove more of the rRNA. Preferably, it would be completely removed or into the single-digit percentile of my sequencing reads.

6.13.1 Further removal of rRNA: strategy

Going from 100% to 80% was a drastic reduction. I think the move from 80% down will be easier. The plan is to run two (or more) samples through the MicrobExpress kit. Then pool the two and run them through a second time. Presumably this second run will contain much less rRNA so the sample loss will likewise be much less.

6.13.2 Further removal of rRNA: first try

Mon Jul 2, 2007

I'm growing up 6 samples in LB. The samples were grown in 5 ml of LB shaking at 300 rpm from a 1:100 dilution. Cells were grown to an OD600 of around 0.6.

Tues Jul 3, 2007

I followed the protocol *Preparation of PET libraries* in the appendix with the following modications: I put 1 μl ready lyse in 500 μl TE. I used 100 μl of this for the lysis rather than measuring the dry lysozyme. I didn't take a sample at SAMPLE POINT A, but I did after SAMPLE POINT B (total volume was 35 μl here):

Sample	DNA (ng/ul)	260/280	260/230	total yield
sample A	1770.0	2.08	2.28	62.0 μg
sample B	1225.0	2.09	2.31	43.0 μg

At SAMPLE POINT B, I also saved 1 μl of each sample to run on a gel (see the first two lanes in Figure 6.13.2).

After SAMPLE POINT B, I took 3 samples of 10 μg from sample A and sample B (6 samples total). One of the three samples was run through the MICROBExpress kit normally. The other two were run through as well, but before the final EtOH precipitation, I pooled the two into a single tube and ran an Isopropanol precipitation rather than an EtOH precipitation. I then took the pooled samples and ran them through again the rRNA removal step a second time. I eluted all of the samples into 16 μl TE [Ambion]. I also measured the RNA concentrations using the Qubit, which is probably more accurate, since it should be less influenced by salts and other junk that have made their way into my RNA over these many steps.

Sample	RNA (ng/μl)	Qubit (ng/μl)	260/280	260/230	yield
A	105.0	74.6	2.12	1.80	1.68 μg
A.next (pooled 1st enrichment)	111.9	-	2.18	2.05	1.79 μg
B	103.4	62.2	2.09	1.83	1.65 μg
B.next (pooled 1st enrichment)	119.2	-	2.15	2.06	1.91 μg
A.2x (pooled 2nd enrichment)	92.6	67.0	2.09	1.59	1.48 μg
B.2x (pooled 2nd enrichment)	76.2	54.8	2.28	1.61	1.22 μg

I saved 2.5 μl of A, B, A.2x, and B.2x for a gel (see the last four lanes of Figure 6.13.2).

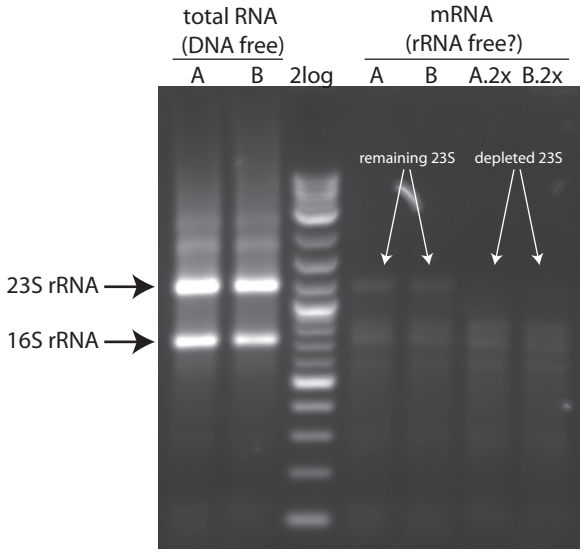


Figure 6.62: total RNA was run through the MICROBExpress kit one time and two times (2x) to remove the rRNA. this is a 1.5% TAE agarose gel

Brief Conclusions: From the gel (Figure 6.13.2), it looks like the rRNA was more depleted this time than in an earlier attempt (Figure 6.12). Actually even only doing 1x rRNA removal looks better this time, but there is a noticable amount of 23S left. After 2x the 23S on the gel looks completely eliminated. It's hard to tell about the 16S, since there might mRNA at that length too. 16S seemed to be pretty much removed last time I used the MICROBExpress kit. We'll know better after sequencing.

Wed Jul 4, 2007

For the 1st strand synthesis, I used 11 μl of each of the four samples (A, B, A.2x, B.2x)+ 1.5 μl of SuperScript II (I ran out of SuperScript III, which is why I used SuperScript II). Unfortunately, I screwed up and didn't follow the superscript manual exactly. I added the buffer before heating to 65C rather than after. I followed the SuperScript II instructions not the Superscript III instructions (e.g. I incubated my random hexamers at RT for 10 minutes rather than 5).

After the 2nd strand synthesis and end-repair, I cleaned up with a Qiagen PCR purification rather than phenol:chloroform. Since I know have the Qubit for measuring low DNA quantities (and also importantly, quantifying DNA in the presence of RNA), I quantified the amount of cDNA using the HS dsDNA reagent:

Sample	**Amount (ng/μl)**	**yield (ng)**
A	17.12	513.6
B	11.48	344.4
A.2x	14.74	442.2
B.2x	14.54	436.2

Because, I could quantify the DNA, I realized I've been using too much adaptor for the adaptor ligation. This time I used only 1 μl of BamISH adaptor (2.1 μg) rather the previous 2 μl .

Thur Jul 5, 2007

I ran the end-repaired cDNA on a TAE Sybr Safe 1% gel for 20 minutes for size-selection. I purposely ran it this short amount of time, so I wouldn't have to cut to large a chunk of gel to cleanup. I cut from $>= 300bp$ for each of the four samples. I cleaned up the samples using a Qiagen Gel cleanup column, eluting into 30 μl . I didn't quantify the DNA, instead I assumed (based on some experiments I did right before this experiment) that the loss from the kit would be about 50%.

Fri Jul 6, 2007

Circularization

I circularized all for cDNA samples (a, a.2x, b, b.2x) using 10 μl of the 30 μl cDNA from the gel extraction above. I used 0.5 μl of circularizer DNA (50 ng), 2 μl T4 ligase buffer, 1 μl T4 ligase, and 6.5 μl H_2O . I used the original dsDNA circularize for samples a and a.2x. I used circularizer 2 for samples b and b.2x. The ligation was for 2hr at 16C, deactivated for 10 min at 65C.

remove linear DNA

I removed the linear DNA with 2 μl of Exonuclease I and 0.4 μl of Exonuclease III for 45 min at 37 C. The reaction was heat deactivated at 80 C for 20 minutes. I did NOT clean up the reaction.

RCA

I ran the following RCA rxn to amplify my circularized cDNA: 5.25 μl 10x RCA buffer [epicenter], 4 μl dNTP, 5 μl template (i.e. uncleaned circular DNA from above), 333.25 μl H_2O , 2.5 μl RCA

hexamer. I heated it to 95C for 5 minutes, cooled it back to RT and added 2.5 μl ϕ29 polymerase [epicenter]. The rxn was incubated at 30C for 12hr followed by 10 min of deactivation at 65C.

Sun Jul 8, 2007

EtOH the RCA

I EtOH precipitated the 4 RCA rxns; I didn't do a phenol:chloroform extraction. I resuspended the nice bulky white DNA pellets in 60 μl of TE. I was trying to get the DNA a little more concentrated than before.

The yields were measured with the nanodrop:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**total yield**
sample a	2022.1	1.80	2.31	121 μg
sample a.2x	2261.1	1.81	2.31	136 μg
sample b	2160.7	1.80	2.30	130 μg
sample b.2x	2374.8	1.81	2.32	142 μg

cut/gel the RCA to make/see PETs

I cut 10 μl (appx 20 μg) of each RCA sample with 7.5 μl MmeI using 1.25 μl of 1:10 SAM, 2.5 μl of buffer 4, and 3.75 μl of H_2O . After starting the rxn, I realized that the glycerol concentration was probably way too high, since over a forth of my rxn was enzyme (and enzymes come in glycerol). A quick chat with Ilaria from my lab and the NEB website, and I had two more verifications that this was a bad idea. Nonetheless I kept going assuming that it would cut at least to some extent.

I ran a 3.5% nusieve TAE sybrsafe gel 6.63. I ran at 120V for 1hr. I cut out all four bands for gel extraction (this is why the image doesn't look so nice. I just used the lab's cheap digital camera, so I didn't have to use UV). I cleaned up all four gel slices with a Qiagen column-based gel cleanup kit.

Brief Conclusions: A and A.2x look fine. Looks like something might be funky with B and B.2x. Maybe I just ran the gel too hot and they diffused away? Next time I'll try less voltage. I also need to remember to use a 25 bp ladder next time in addition to the 2-log.

cloning the PETs

Mon Jul 9, 2007

I end-repaired the gel-cleaned PETs [Epicenter] for samples A and A.2x. I cut 2 μg of pUC19 with SmaI for 30 min at RT followed by cleaning up with a Qiagen PCR cleanup kit. I eluted into 30 μl EB. Yield was 41.0 ng /μl , 1.86 (280/260), 2.05 (230/260). I made a 15 μl dephosphorylation rxn with 1.5 μl antarctic phosphatase buffer, 1 μl antarctic phosphatase, 190 ng (4.6 μl) cut pUC19, and 7.9 μl of H_2O . I ran this reaction at 37C for 60 min followed by a 5 min deactivation at 65C. I ligated as follows: 2 μl T4 buffer, 1 μl high concentration ligase, 2 μl dephos vector (appx 25 ng), 10 μl insert, 5 μl H_2O . I ran the rxn for 15 min at RT followed by 65C for 10 min. I transformed 2 μl of the ligation into DH5α. I plated 100 μl of each tranformation onto an amp plate with X-gal.

Unfortunately, the plates were *filled* with blue colonies. I think either the smaI or the antarctic phosphatase aren't working anymore. This also explains why Steve in the lab (and Esther too) couldn't get their blunt clones to work. Both of the enzymes were really old. I bought new ones. And I'll try again.

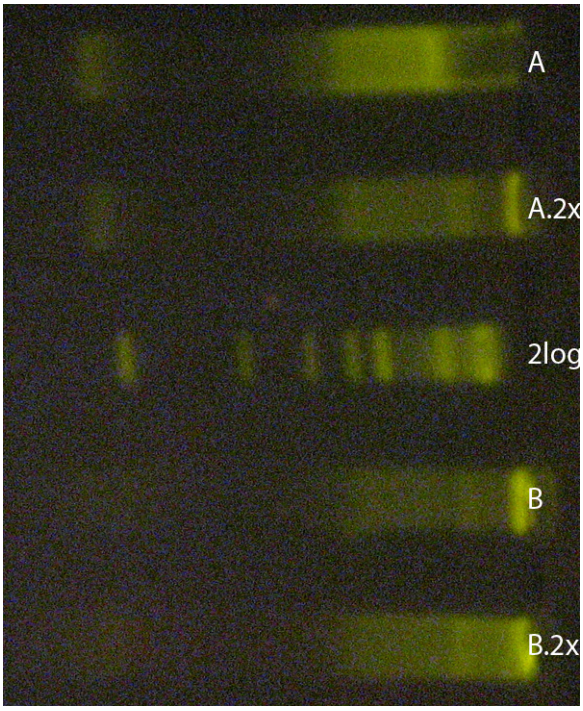


Figure 6.63: cut PETs from a, a.2x, b, b.2x imaged on a transilluminator

Improving the yield of digested PET

Tues Jul 10, 2007

I want to mess around a little to try and get more PET from my RCA. Am I using too much MmeI, too little? Too much DNA? Too little?

Unlike last time, I'm going to use a much higher volume rxn to prevent excess glycerol in my rxn. I'm trying four 100 μl digestions. All four contain 10 μl of NEB4 plus 1.6 μl of 1:10 SAM. I used RCA sample A from above (2.02 μg /μl). All rxns were incubated 1hr at 37C.

digestion	**RCA (μl)**	**MmeI (μl)**	**H_2O (μl)**
A	5	3.75	79.65
B	5	5.00	78.4
C	5	5.00	73.4
D	10	10	68.4

I cleaned up the reactions with a Qiagen PCR kit and eluted into 30 μl of EB buffer. I added 5 μl of 1:20 fisher dye and loaded them into a 4% TAE 60 ml agarose gel with SybrSafe (Figure 6.64). I also ran one well with NEB 2log ladder + 4 μl of full strength fisher dye to determine the approximate migration times of the three dyes for future reference. I ran the gel for 80 minutes at 110 V, which placed the red dye at the end of the gel. In hindsight, I think 110V was a little hot; go with 100 V next time.

One big problem with this experiment was that my DNA seemed to float/diffuse upwards when I loaded it into the well. This happens sometimes when I have DNA cleaned with Qiagen PCR cleanup kits. What causes this? Residual ethanol? A lack of salt? A lack of EDTA? It is highly annoying to watch your hard-earned sample float away, so I investigated this a little further to hopefully prevent this in the future. See section 10.7 on page 402 for details.

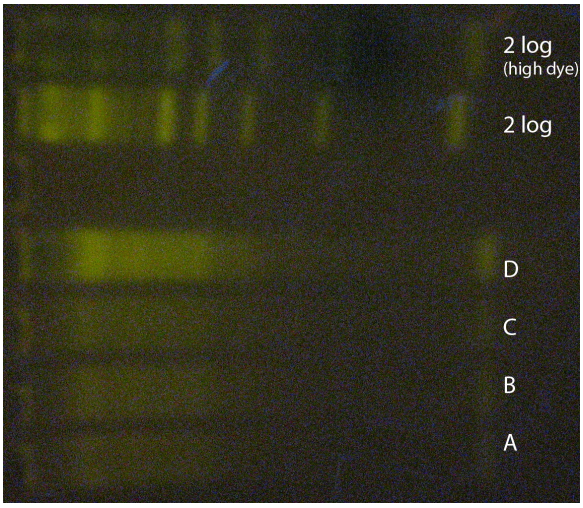


Figure 6.64: PETs from cuts A, B, C, and D imaged on a transilluminator

Brief Conclusions: Unfortunately, 5 μl of RCA (10 μg) doesn't seem like enough to be able to detect a different in intensity between my bands. 10 μl of RCA (20 μg) was certainly better, and I should use that as my starting point in the future. If anything, I'd say that the lesser quantities of mmeI still cut well.

improving PET digestion yield using more RCA DNA Wed Jul 11, 2007

I tried two 100 μl rxns with 10 μl (20 μg) of RCA using 10 μl and 5 μl of MmeI. I also tried one 200 μl rxn using 20 μl RCA and 20 μl enzyme. Digestions were for 1hr at 37C. Instead of using the Qiagen PCR purification kit, I used a microcon YM-30 to concentrate the digestion and hopefully avoid the problem of having my sample diffuse away after I loaded it. Plus with the microcon, you can get lower volumes (10 μl), which would allow me to use less agarose (cheaper, better picture, easier to gel extract). I ran the gel at 90V until the purple dye was at the very edge to try and get even better separation than before. The gel was junk and I didn't bother taking a picture. Two things could've happened (and I think it was a mixture of the two): 1) the microcon didn't recover a big percentage of my sample; 2) I ran the gel so long that the small DNA started to diffuse too much.

Brief Conclusions: Next time: try minElute from Qiagen, elute with TE. Add a little salt to the eluted DNA in TE? run gel for less time

6.13.3 further removal of rRNA: 2nd try focus on ESTs

Fri Jul 20 19:58:35 EDT 2007

I'm in a bit of a time crunch trying to prepare some slides for a talk. I don't want to mess with cloning the blunt PETs when I can know straight away if the MICROBExpress 2x trick worked with the easy to sequence ESTs. So I'm going to make a couple new samples and attach a BamHI adaptor instead of a BamISH adaptor. I'll then clone and sequence a few clones to see the frequency of rRNA.

growing the cells

Fri Jul 20, 2007

I grew up two samples labeled *e* and *f* from a 1:100 dilution from an overnite that Ilaria made for me (I was still in West Palm Beach, FL). The cells were grown at 300 rpm in LB for appx 3 hr to an OD of 0.692 and 0.645 (background LB subtracted) for samples e and f respectively. 2.5 ml of this culture as added to 5 ml of RNAprotect [Qiagen], vortexed 5 sec, incubated at RT for 5 min, and spun at 4000 rpm in a bucket centrifuge for 15 minutes. The pellets were placed at -20C.

RNA preparation

Sat Jul 21 21:13:33 EDT 2007

RNA was prepared as per my PET protocol in the appendix but with the following modifications. Like the previous RNA prep, I used 1 μl of ReadyLyse [Epicenter] in 500 μl of TE rather than weighing out the power lysozyme. I didn't take a SAMPLE A. I spec'd the RNA at SAMPLE POINT B:

Sample	DNA (ng/ul)	260/280	260/230	total yield
sample E	1826.6	2.07	2.25	63.9 μg
sample F	1676.7	2.05	2.25	58.5 μg

I took 0.5 μl of each sample to run on a gel.

I used 5.5 μl and 6.05 μl of sample E and sample F respectively (appx 10 μg) for the MICROBExpress rRNA removal. As in my previous attempt at 2x rRNA removal, I ran 3 rRNA removal rxns for each sample. Unlike the previous time, I did not combine the two samples to be used for a second round of MICROBExpress, as I felt like the isopropanol precipitation caused me to lose some of my sample relative to the EtOH precipitated samples. I eluted the normal samples in 15 μl of TE. I eluted the samples to becombined into 10 μl each (20 μl total combined RNA).

Sample	RNA (ng/μl)	Qubit (ng/μl)	260/280	260/230	yield
E	123.2	102	2.17	1.98	1.85 μg
E.next (pooled 1st enrichment)	166.4	134	2.16	2.06	3.33 μg
F	104.0	88.2	2.16	2.06	1.56 μg
F.next (pooled 1st enrichment)	150.9	120	2.18	2.12	3.02 μg
E.2x (pooled 2nd enrichment)	118.2	95.4	2.12	1.75	1.89 μg
F.2x (pooled 2nd enrichment)	108.9	86.6	2.13	1.74	1.74 μg

Brief Conclusions: The yields for the pooled 1st enrichment are much closer to what I'd expect for pooling two samples together. They are much closer to having two times the total yield of the unpooled samples. Compare this table with the one in section 6.13.2 on page 321.

Sun Jul 22, 2007

I used 11 μl of all of the samples for 1st strand synthesis of cDNA (E, F, E.2x, F.2x), which is approximately 1.5 μg of each. My Superscript III arrived, so I followed the standard protocol this time (incubate 50C 1hr). I used 1.5 μl of Superscript III.

The yields of ds-cDNA prior to adaptoring (as measured by Qubit HsDNA) were:

Sample	Qubit (ng/μl)	yield (ng)
E	30.4	912
E.2x	20.8	624
F	27.8	834
F.2x	15.46	463.8

I'm using BamHI adaptor NOT BamISH adaptor, since I plan to clone into pUC19. I performed the blunt adaptor ligations with BamHI adaptor just like I did with the BahISH adaptor in the previous MICROBExpress 2x attempt.

Mon Jul 23, 2007

I ran a gel of the total RNA and mRNA sample (Figure 6.65); everything looks normal enough.

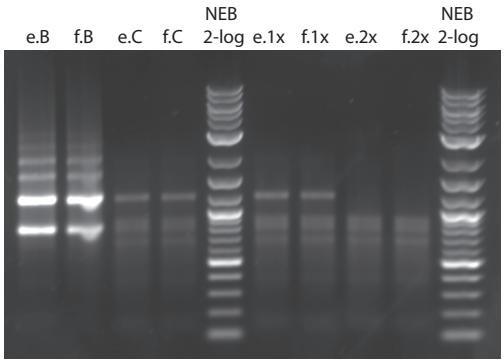


Figure 6.65: Total RNA (e.B, f.B), 1x MICROBExpressB (e.C, f.C, e.1x, f.1x), 2x MICROBExpress (e.2x, f.2x). Note that e.2x and f.2x are derived from e.1x and f.1x (i.e. after running e.1x and f.2x through the kit a second time, I had enough 1x left over to throw some on the gel and compare to the 2x samples).

I also ran two gels to size-select the cDNA prior to cloning. For e and e.2x, I loaded the samples directly onto a wide-comb (6-well) 45 ml 1% gel and ran it 20 min at 90V. For f and f.2x, I first ran the samples through a PCR purification to try and remove some of the adaptor (didn't really help there was still tons of adaptor on the gel). I ran this gel in the exact same way as the e and e.2x gel except I used a 10-well comb and a 65 ml gel. Both gels were SybrSAFE. Next time I should run the gels for 25-30 minute to help increase the separation away from the adaptor; after 20 minutes they're still pretty close.

To Do!!! It would be nice to know if you can aid in the removal of the short adaptor relative to the longer cDNA by running the PE buffer across the sample multiple times during the Qiagen PCR purification process.

To clone the adaptored cDNA, I cut 2 μg of pUC19 using the following rxn: 2 μl pUC19, 2 μl 10x BSA, 2 μl NEB3, 13 μl H_2O , 1 μl BamHI enzyme. I ran the rxn for 45 min at 37C. I cleaned up the reaction with a Qiagen PCR purification kit. I ran two reactions because the first one a screwed up and used NEB2 rather the optimum NEB3, normally I wouldn't care, but since I'm only cloning using one-cutter, I want to make sure I cut this plasmid well. The yield of cut DNA was:

Sample	DNA (ng/ul)	260/280	260/230	total yield
pUC19 NEB2	59.3	1.89	2.29	1.78 μg
pUC19 NEB3	62.9	1.87	2.24	1.89 μg

I ran a phosphotase reaction to prevent plasmid self-ligation: 10 μl cut pUC19 (the NEB3 sample), 7 μl H_2O , 2 μl antarctic phosphatase buffer, 1 μl antartartic phosphatase.

Finally, I prepared the ligation of the cut, dephosphorylated buffer with the adaptored phosphorylated size-selected cDNA: 2 μl T4 ligase buffer, 10 μl cDNA gel purified, 2 μl dephosphorylated pUC19, 5 μl H_2O , 1 μl T4 ligase. I ran the rxn 30 min at 16C. I cloned the ligation into DH5α and plated 50 μl of each sample (e, e.2x, f, f.2x).

Tues July 24, 2007

I picked 24 colonies and grew them up in LB. After $>$ 8hrs of growth, I miniprepped all 24 samples.

Wed July 25, 2007

I digested all 24 minipreps with 2 μl EcoRI buffer, 10 μl plasmid, 0.5 μl EchoRI, 0.5 μl HindIII, 7 μl H_2O . The rxns were run 15 minutes at 37C followed by heat deactivation at 65C for 10 minutes. I ran the digestions on an agarose gel (actually two because I didn't have enough lanes). Unfortunately, I goofed a couple different ways. First, I tried using the matrix-impact 2 variable-spacing-electronic-multichannel-pipettor to put the first four samples in at the same time. A nice time saving idea, but I learned that electronic multichannel pipettors are a very bad way to load a gel. The problem is that as soon as you push the dispense button on a multichannel, *all* of the sample is going to come out whether or not you're got the pipettes accurately centered or not. With the manual multichannel pipettor, you can dispense little-by-little – making sure to optimize the amount of sample that falls to the bottom of the gel well as you go along. However, there is no manual multichannel with variable tip spacing. So in short, *I lost the first 4 digestion samples* into the gel buffer. The second problem is that I'm filling this section out in the middle of September and I didn't mark on a sheet of paper the order of the samples on the gel. I'd assume it goes e5, e.2x 1, . . . , f1, . . . , f.2x 1. However, it doesn't matter too much because the gel (Figure 6.66) looks just fine, and virtually all inserts are >500bp, which is what I wanted to see.

Gotchas: Don't load an agarose gel with an electronic multichannel pipettor. Manual multichannels are fine for this purpose (after a little practice).

Since the gel looked fine (Figure 6.66), I spec'd 20 samples (the first five from every sample type) and sent them out for sequencing. The spec values and the gene the sequenced EST best matches on the genome is shown in the table below.

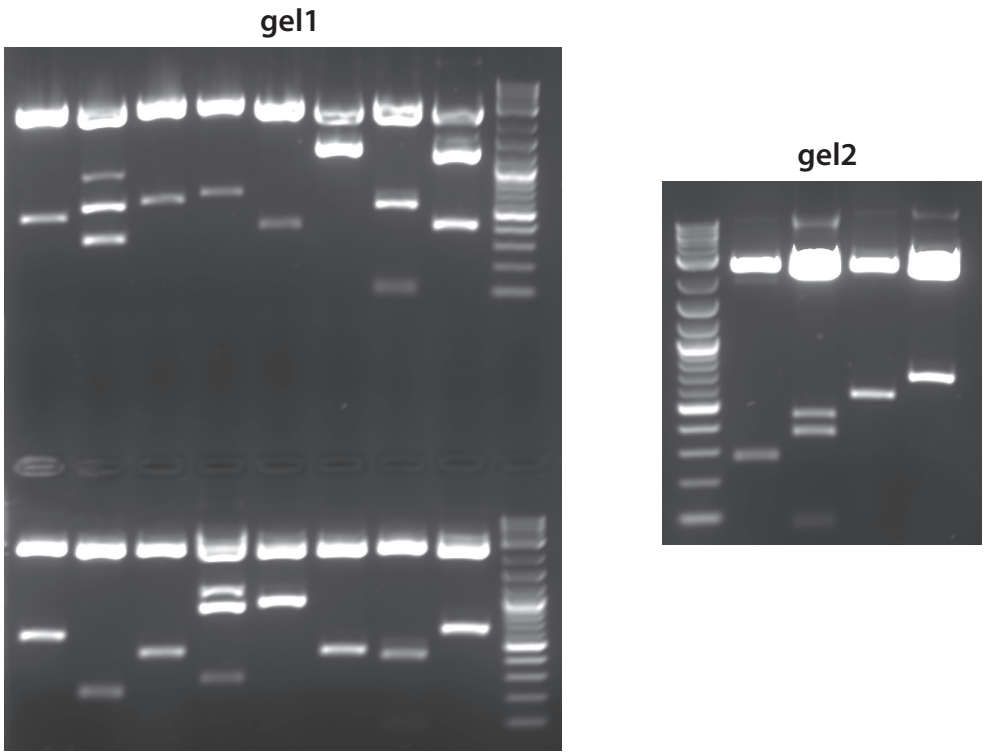


Figure 6.66: Cut ESTs cloned into pUC19 are in general longer than the 500 bp gel selection length.

Sample	Time	ng/ul	260/280	260/230	Top Hit
e1	11:56 AM	370.89	1.96	2.26	16S
e2	11:57 AM	379.76	1.96	2.44	23S
e3	11:59 AM	209.03	1.98	2.25	?
e4	11:59 AM	308.71	1.97	2.24	23S
e5	12:00 PM	415.84	1.94	2.28	23S
e.2x 1	12:01 PM	140.57	2.01	2.46	mdlA
e.2x 2	12:03 PM	349.38	1.96	2.29	23S
e.2x 3	12:03 PM	431.42	1.93	2.25	23S
e.2x 4	12:04 PM	512.19	1.91	2.23	tnaC
e.2x 5	12:05 PM	513.86	1.9	2.22	?
f 1	12:05 PM	302.78	1.95	2.29	23S
f 1	12:06 PM	274.55	1.95	2.27	mglB?
f 3	12:06 PM	298.16	1.95	2.25	23S
f 4	12:07 PM	505.84	1.88	2.19	23S
f 5	12:08 PM	382.68	1.94	2.26	16S
f.2x 1	12:09 PM	290.61	1.98	2.29	16S
f.2x 2	12:10 PM	376.87	1.94	2.24	23S
f.2x 3	12:10 PM	275.03	1.98	2.25	23S
f.2x 4	12:11 PM	530.4	1.89	2.2	?
f.2x 5	12:11 PM	248.21	2.01	2.31	proS

The three questions marks in the table above are the sequencing rxns that failed. Agencourt ran them again and they worked. However all three mapped back to rRNA (can't remember if it was 16S or 23S, but who cares, I hate them both).

To Do!!! I just realized with the MicrobeExpressB kit there might be a faster way to cycle the procedure. I currently run through the kit, concentrate with EtOH, and start over. The EtOH precipitation step takes up the majority of the time. Rather than cleaning up, I can just take the elution (ignor the wash step, hopefully won't result in too much loss) and start over again with new oligo and beads. Clearly runing 2x through the kit is helping. I'd guess I'll get dimminishing returns in terms of rRNA removal as I go through 3x, 4x, ..., but it is worth a shot. Another strategy, use some 1st strand cDNA to run a PCR reaction to amplify full length rRNA (get general primers to all *E. coli* rRNA [is all *E. coli* rRNA sequence exactly the same?]). During the amplification, use a biotinylated primer for the reverse strand. Now catch the biotin on a dynal bead and strip away the other strand. Whenever you finish with other first strand reactions in the future, add the biotinylated rRNA plus a dsDNA exonuclease to zap the first strand cDNA complement to the rRNA. This should work if the dsDNA exonuclease doesn't also attack RNA:DNA hybrids. If it does attach hybrids, I'd need to do a RNAse digestion, clean up the cDNA, and then use the exonuclease trick.

results and conclusions from the further removal of rRNA focus on ESTs

Tue Sep 18 18:57:34 EDT 2007

Unfortunately, single-read Sanger DNA sequencing is still kinda expensive to be running hundreds of test samples. So I'm going to just calculate some crude stats with the 20 or so samples that I do have.

For this round alone, I ran 10 samples with 1x MicrobeExpressB and 10 samples with 2x MicrobeExpressB. As I found out many months ago, 1x MicrobeExpressB removes a drastic amount of rRNA (I've was never able to sequence mRNA from a sample that was *not* run through MicrobeExpressB, it has all been rRNA). From the gel alone it appears that the second round removes addition amounts of rRNA (see Figures 6.13.2 and 6.65). In the table above 1 in 10 (10%) samples in the 1x MicrobeExpressB was mRNA, while 3 in 10 (30%) samples were mRNA for the 2x MicrobeExpressB. Seventy-percent unwanted sequence still sucks, but it is a hell-of-a-lot better than ninety-percent. I'd like to get at least 50% mRNA. Hopefully the MicrobeExpress cycling idea or the dsDNA exonuclease idea will give me that extra boost.

With my previous results with 1x MicrobeExpressB, the ratio was quite a bit better. The overall ratio for 1x MicrobeExpress is 3 in 20 (15%) mRNA.

One final thing to note, tnaA has now been cloned two independent times. Given that I've only sequenced six mRNAs so far (and lots of rRNA), it's kinda odd that one gene has shown up 2x already.

6.14 How many samples must be pooled for one 454 run

After all the work trying to recircularize the PETs (or do some sorta single-primer PCR), I've decided to consider brute force accumulation of PETs. It's clear that I don't get a huge amount of PET DNA when I run the MmeI cut the RCA amplified cDNA (see for example Figure 6.44). However, I want to pool the samples anyways, because I don't want to just run one species, one condition in my pyrosequencing reaction.

So Tim suggested I forget all of the fancy 2nd amplification strategies, and just pool a bunch of samples together. The question then is: *how much DNA do I get from one MmeI digestion and PET purification.*

6.14.1 Amplifying the PET DNA

Wed, Sep 19, 2007

I ran the following RCA reaction on samples A.1x and A.2x from section 6.13.2 on page 320. 5.25 μl 10x RCA buffer, 2 μl dNTP, 5 μl template, 2.5 μl hexamer, 35.25 μl H_2O . Incubate at 95C for 5 minutes, place on ice, add 2.5 μl ϕ29 enzyme and incubate at 30C for 12 hr followed by heat inactivation at 65C for 10 minutes.

Thur, Sep 20, 2007

I concentrated the RCA reaction with EtOH and resuspended in 60 μl TE. The yields from these reactions were:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**total yield**
RCA sample A.1x	898.5	1.80	2.38	53.9 μg
RCA sample A.2x	995.2	1.80	2.38	59.7 μg

Brief Conclusions: The yields were half what I got when I performed the exact same reaction back on July 4 (see the table in section 6.13.2 on page 323). Maybe something is going bad with my DNA, enzymes, or dNTPs?

Fri, Sep 21, 2007

I cut approximately 40 μg of RCA samples A.1x and A.2x from above. For each digestion, I used 40 μl RCA DNA, 30 μl MmeI, 14.1 μl SAM, 45 μl NEB4, 320.9 μl H_2O for a total volume of 450 μl . I ran the reaction at 37C for 30 minutes in a water bath. I then precipitated each reaction in EtOH plus 1 μl of glycoblue. I resuspended the cut DNA into 20 μl of TE. I ran all 20 μl of the two samples (plus an extra 5 μl of loading dye) out on a Nusieve 4% agarose gel (Figure 6.67). The gel was SYBR Safe, I used 1:20 fisher dye, I ran the gel 45 min at 90V and 45 min at 120V. The gel looked great by eye. And although not perfect when photographed by a point-and-shoot camera, it wasn't too bad either.

I cut extremely thin slices around the PETs A.1x and A.2x. For controls, I cut the 125 bp fragment from the Invitrogen ladder, and I cut another band at around 100 bp in a lane with no DNA as a negative control. I cleaned the reactions up with a Qiagen column-based gel purification kit, and I eluted into 30 μl of EB. I quantified the DNA using 20 μl of the purified DNA with the hsDNA Qubit kit.

Sample	**DNA (ng/ul)**	**total yield (ng)**
gel purified PET A.1x	0.538	16.14
gel purified PET A.2x	0.576	17.28
125 bp band from ladder	1.956	58.68
gel slice from DNA-free lane	0.0254	0.762

Brief Conclusions: Well, the yields weren't so great, but they're much better than background. The empty gel-slice resulted in 0.762 ng yield of DNA (presumably this is just background noise from the Qubit), while the purified PET gel-slices resulted yields of around 20 ng. I'm not sure

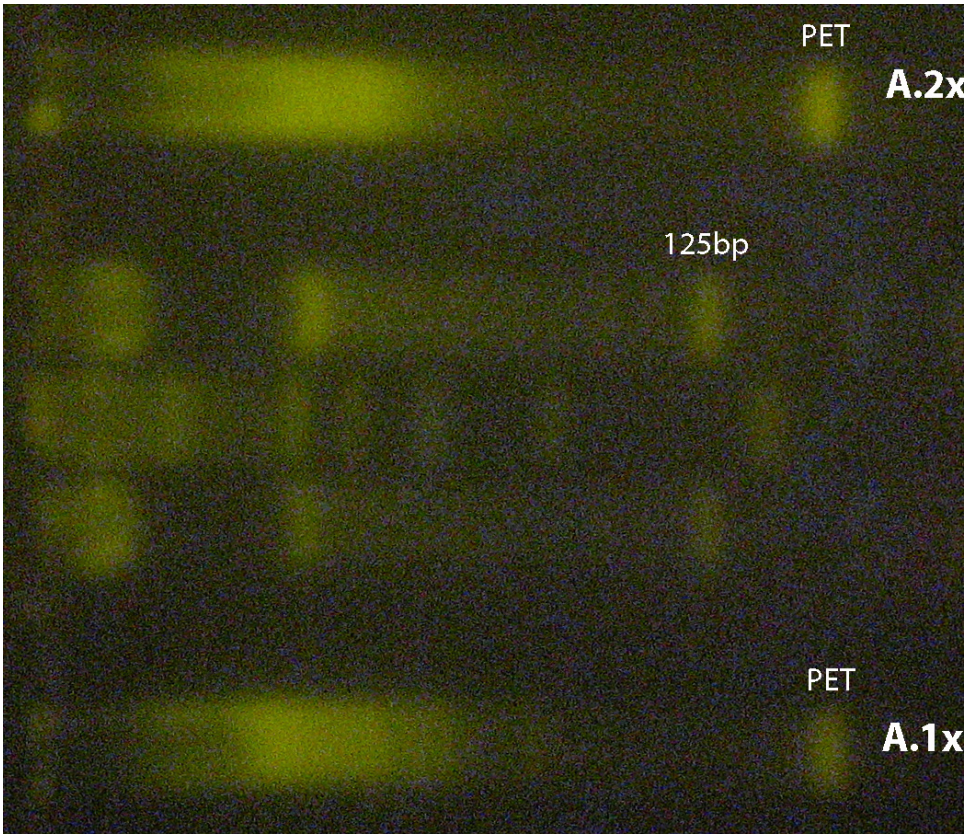


Figure 6.67: The lanes from top to bottom are MmeI cut A.2x, bank, 4 μl 25 bp ladder (invitrogen), NEB 2-log ladder, 2 μl 25 bp ladder (invitrogen), MmeI cut A.1x.

why the 125 bp piece had such a high yield, because on the gel it actually looks weaker than the PET (Figure 6.67). Maybe the PET is a little short, so it doesn't purify as well? Maybe I was too conservative when I cut the band out (most likely?)[27]. As things stand, I need 2 μg of DNA to give to the pyrosequencing folks. That would require pooling 100 samples (ouch!). However, I could cut more RCA, since I still have quite a bit left from each RCA run. Perhaps, I need to try other gel purification methods or optimize my digestion to use less MmeI, or this will cost quite a bit per sequencing run on MmeI enzyme.

[27] I did purposely cut a tiny-bit inside the PET band to lessen the amount of non-PET DNA, maybe the Qubit dye is more sensitive to longer DNA?

Chapter 7

Removal of rRNA from prokaryotic total RNA samples

THIS CHAPTER/PROJECT IS IN PROGRESS

In Chapter 6, I am trying to generate and sequence paired-end-tags from prokaryotic mRNA. All companies that offer highly-parallel sequencing machines are also working to develop mechanisms for paired-end-reads. Presumably, with their large crews of folks working on this problem, someone is going to figure out a good solution, which will then make my solution (based on the work of Shendure *et.al.* Science 2006) in chapter 6 obsolete.

However, one theme that has begun to dominate chapter 6 and which is a more general problem for all prokaryotes that will continue even after the development of efficient paired-reads technologies is – how to remove rRNA from total RNA in prokaryotes. I'm going to test a few different ideas related to the rRNA removal problem in this chapter.

7.1 developing a gentle, quick-lyse procedure that produces undegraded total RNA

Several of the ideas I have (or that I've received from others) for removing rRNA from total RNA require a fast gentle total RNA prep procedure that does *not* use harsh chemicals (e.g. those normally used to inhibit RNAses also mess up protein folding in general) and does not result in degraded RNA despite the absence of typical RNAse inhibitors. Not a small order...

I'm going to try and develop such a procedure in this section. I'll assay RNA quality by looking for the standard 23S and 16S bands on an agarose gel.

7.1.1 quick-and-simple readyLyse preparation of total RNA try 1

Dec 5, 2007

growing the cells

I added a 1:50 dilution of MG1655 overnite culture into 4 ml of LB in a 12 ml falcon tube. I grew the cells for 3hr to an OD600 of 0.8 (a little higher than I wanted).

lysing the cells

I took 1.5 ml of the OD600 0.8 culture and resuspended it into 15 μl TES (TE + 100 mM NaCl). I tried both normal and 10x ready-lyse [Epicenter] amounts (29U and 290U total). The samples were incubated for 5 minutes at RT. The samples were spun down at 13K rpm for 1 minute and the supernatant was retained as a RNA/genomic DNA mixture. 1.5 μl of Turbo-DNA-free [Ambion] was added and the samples were incubated at RT for 3 minutes to degrade the genomic DNA.

running on a gel

I ran the entire 15 μl on a 1% TAE gel with NEB ssRNA sample buffer (Figure 7.2 lanes labeled 1 and 2 are the 1x and 10x ready-lyse samples respectively).

Brief Conclusions: Rubbish. Not sure what that is on the gel (Figure 7.2). There is certainly nucleic acid material there, but it is pretty short to be genomic DNA and doesn't have the characteristic 16S and 23S rRNA bands I'd expect to see if it were total RNA. Next time I need to add DNAse before I run the sample on a gel and to try using supernase to inhibit degradation in case what I see here is just degraded RNA.

7.1.2 quick-and-simple readyLyse preparation of total RNA try 2

Dec 12, 2007

I want to retry the quick-lyse procedure to see if I can get a decent rRNA band. First I'm going to use superase [Ambion] RNAse inhibitor. Second I'm going to try using very low levels of formaldehyde. The formaldehyde might deactivate RNAses and perhaps it will crosslink the rRNA-ribosome complex together and make it easier to spin down for removal?

I grew an 1:50 dilution of overnite MG1655 culture in 4 ml LB in a 12 ml falcon tube for 2 hr and 37 minutes the OD600 samples for the three tubes were 0.734, 0.722, 0.763 (not bkgrd subtracted). I took 1.5 ml of culture for each RNA sample. Sample 2 was mixed with 0.1% formaldehyde for 10 minutes at RT and quenched with glycine to stop the crosslinking reaction. I used 1x lysozyme and TS media (10 mM Tris, 100 mM NaCl) instead of TES in case the EDTA was inhibiting the RNAse H before. I washed all cells 1x in 100 μl TS. I resuspended them in 25 μl TS + 1 μl superase (20 Units for samples 1-3; sample 4 had no superase). I added 1 μl of 1x lysozyme and lysed for 15 minutes at RT (vortexing briefly every 3 minutes). I spun the lysed samples for 5 minuts at 13K rpm and keep the supernatant. I added 1 μl of DNA-free Turbo DNAse [Ambion] and incubated for 15 min to remove DNA from the samples. I then added 1 μl RNAse H to sample 3 and incubated for 10 minutes at 37 C.

I ran all four samples on a 1% TBE gel for 50 min at 120 V (Figure 7.1).

Brief Conclusions: More rubbish... As Figure 7.1 shows (lanes 1-4) my quick lyse procedure sucks. I'm about to jump ship on this idea. I need the quick, gentle lyse procedure if I want to try to remove rRNA via ultracentrifugation, but the lyse procedure seems a long way off, since I really haven't moved forwards with any ideas that worked so far. I'm particularly disappointed that the formaldehyde sample didn't work. I figured that would deactivate the RNAses and allow me to have relatively clean total RNA with a 16S and a 23S band. Perhaps my problem is not RNAses?

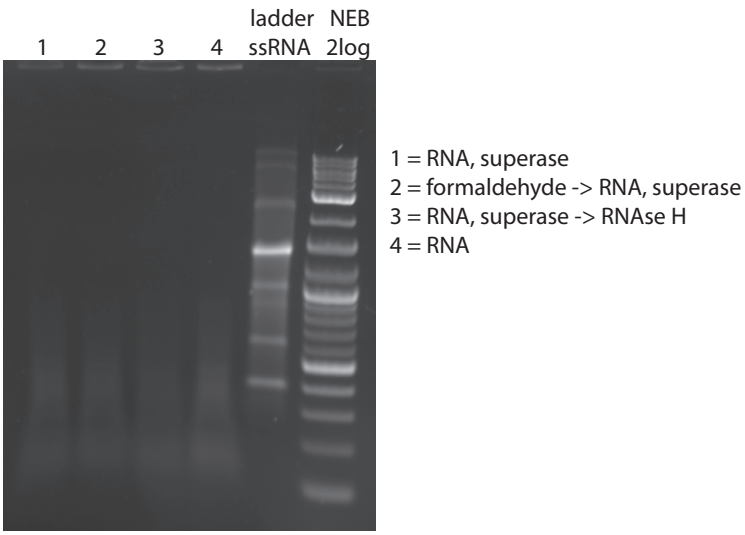


Figure 7.1:

7.2 problems with assaying rRNA removal

Tue Dec 11 17:12:21 EST 2007

On of the challenges with removing rRNA is that the rRNA bands are the most common way to judge the quality of the RNA purification procedures. The distinct 16S and 23S bands are the hallmark of a good RNA prep. So when I attempt to remove those two bands, it is difficult to determine if I've remove those two bands or if I just degraded those two bands into a smear of rRNA. One way to better judge this rRNA removal is with a ssRNA ladder. I have some ssRNA ladder from NEB (Part No: N0362S) to use for this purpose.

7.3 rRNA removal via RNAseH

Write strategy here.

primers

I designed the 23S primers using an alignment of *E. coli* 23S sequences and primer3. The 16S primers are universal 16S primers from the supplemental material of Gill *et.al.* Science 2006.

```
----- 16S primers -----
Bact-8F
5' AGAGTTTGATCCTGGCTCAG
Bact-1510R
5' CGGTTACCTTGTTACGACTT

----- 23S primers -----
forward
5' GACTAAGCGTACACGGTGGAT
reverse
5' TTAAGCCTCACGGTTCATTAG
```

Note that the Bact-8F primer is NOT an exact match to *E. coli* 16S rRNA

```
 Score = 32.2 bits (16), Expect = 0.007
 Identities = 19/20 (95%)
 Strand = Plus / Plus


Query: 1       agagtttgatcctggctcag 20
                |||||||||||  ||||||||
Sbjct: 4206177 agagtttgatcatggctcag 4206196
```

I ordered both primers (the mismatch and the exact match primer).

7.3.1 RNAseH tests

Dec 5, 2007

As a first step to removing the rRNA using RNAseH + PCR amplicon (DNA) of rRNA, I took some total RNA from the -80C (sample 1 from the table in section 6.12.2 on page 304 and some ssRNA ladder [NEB]. I wanted to first check that the RNAse H would not degrade the RNA in the absense of DNA (NEB claims their RNAse H does not contain additional contaminating RNAses). I also wanted to see if I could degrade the RNA with RNAse H by adding a large amount of genomic DNA (which would presumably bind to the mRNA in my sample. I would not likely have enough genomic DNA to remove the rRNA bands, but perhaps everything except the rRNA bands would be degraded). I also assumed the RNA ladder would also not be degraded by RNAse H when genomic DNA was added for similar reasons to the rRNA explanation above. I used 1 μg of RNA ladder, 1 μg of total RNA, and 1 μg of genomic DNA sample 1 from section 5.2.4 on page 214. The RNAse H degradation was for 30 min at 37C.

Since the RNA and genomic DNA samples were pretty old, I spec'd them again with the Nanodrop:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**
total RNA	2235.2	2.04	2.26
genomic 1	409.2	2.14	2.04

The samples were run on a 1% agarose gel (Figure 7.2) for 35 minutes at 110 V.

Brief Conclusions: All the samples are such a blurry mess, its hard to say anything about this experiment (Figure 7.2). Is my RNA just being degraded? Is the gel not running properly? Why does the - control RNA ladder only sample look so bad (lane L/N/N)?

7.3.2 can I get a decent RNA ladder gel

Dec 6, 2007

Sanity check. Can I at least get a decent looking RNA ladder on an agarose gel? Do I have RNAse contamination somewhere that's really screwing me up? What's goin on?

I put 2 μl (1 μg) of the RNA ladder into TE. The first sample I added ssRNA sample buffer [NEB], heated to 60C for 5 minutes; the second sample was run in standard sucrose agarose loading buffer/dye. I switched to a TBE gel and ran the gel for 40 minutes at 120V (Figure 7.3).

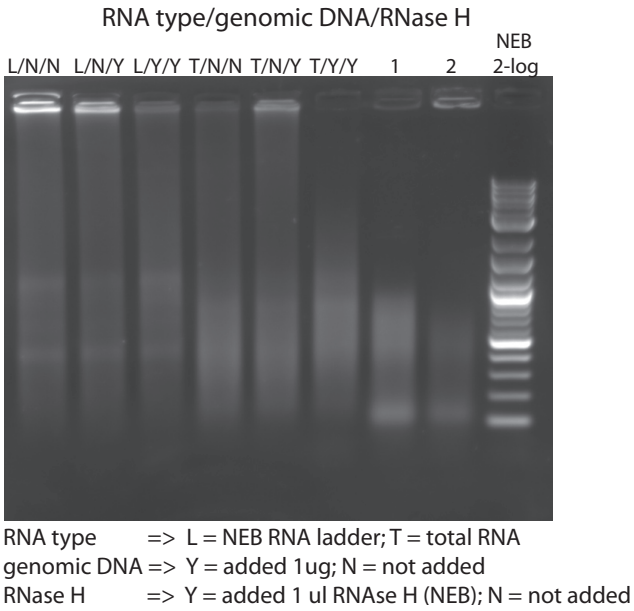


Figure 7.2:

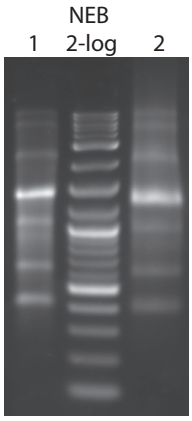


Figure 7.3: Sample 1 contains NEB ssRNA sample buffer and was heated. Sample 2 contains standard sucrose loading buffer.

Brief Conclusions: Certainly a step up. I can at least make out the bands of the ladder (compare Figure 7.3 with the lanes labeled 1 and 2 in the earlier Figure 7.2). Still not terribly good, but the ladder in the image that NEB sends as an example is also quote blury. The sample with the ssRNA sample buffer certainly looks less blurry. I'll continue to use the ssRNA sample buffer and I'll switch to TBE gels.

7.3.3 testing RNAse H on total RNA

Thur Dec 13, 2007

I've heard that RNAse H might have some contaminating RNAses which will degrade my RNA. When I was reading the superase manual, I noticed that superase has a nice feature that it inhibits

most RNAses *except* RNAse H: beautiful. If my RNAse H enzyme has contaminating RNAses I should be able to inhibit them with the addition of superase to my total RNA, yet the RNAse H will still function.

To test this, I'm going to run standard RNAeasy total RNA preps. I'll add superase to some but not others and then test the effect of RNAse H on all of the samples. To add some complementary DNA to my samples (for the RNAse H to use), I'm going to prepare a few of the samples without using DNA removal (via LiCl precipitation).

MG1655 samples were grown 1 hr 50 min from a 1:50 dilution of overnight culture into LB. The OD600 for the two samples was 0.499 and 0.518 (not background subtracted). I used 2.5 ml of culture and 5 ml of RNAprotect. One of the two samples was placed at -20C for 30 minutes with 1/2 volume of 7.5 LiCl to precipitate the RNA (and thus remove the genomic DNA). The second sample was placed at 4C during this time.

Prior to LiCl the samples were eluted into 100 μl of RNAse free H_2O , the yields were:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**yield**
sample 1	1122.5	2.16	2.36	112 μg
sample 2	759.6	2.18	1.93	76 μg

After the LiCl, sample 1 was resuspended into 50 μl of RNAse free H_2O , the yield was:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**yield**
sample 1	1464.3	2.13	2.47	73.2 μg

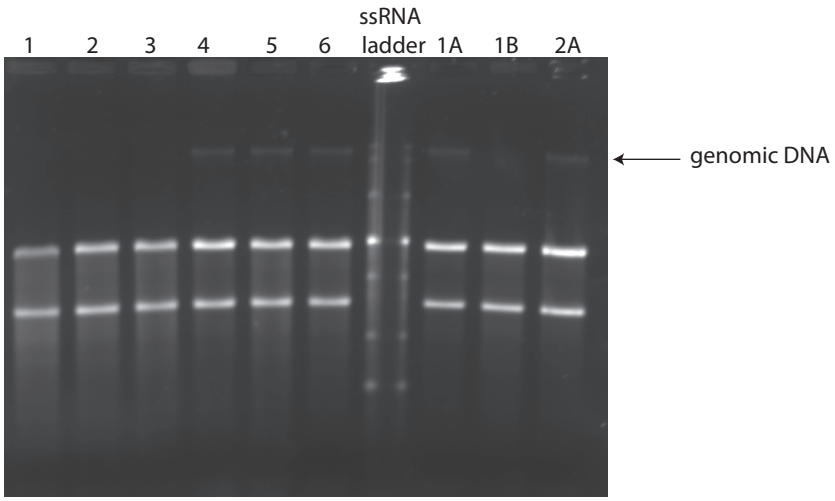
From the above tables, sample 1 was split into 3 and becomes samples 1-3 below. Likewise, sample 2 becomes samples 4-6 below.

sample	**LiCl**	**superase**	**RNAseH**
1	yes	no	yes
2	yes	yes	no
3	yes	yes	yes
4	no	yes	yes
5	no	no	yes
6	no	yes	no

1 μl (20U) of superase was used. 1 μl (5U) of RNAse H was used with an incubation at 37C for 15 min.

I ran approximately 1 μg of each sample on a 1% TBE agarose gel with ssRNA sample loading buffer [NEB] (Figure 7.4).

Brief Conclusions: It looks like the RNAse H from NEB is so free of contaminating RNAses that the cool superase trick is unnecessary (Figure 7.4). Thankfully, both of the preps resulted in very clean total RNA with the characteristic 16S and 23S bands (Figure 7.4 lanes 1B and 2A). It's hard to say if the genomic DNA in samples 4-6 allowed the RNAse H to have any activity. If it did, then we'd expect sample 6 to look different than samples 4 and 5, which doesn't seem to be the case. Overall, not a bad result here though, as it looks like I don't have to worry about my RNAse H containing contaminating RNAses that will degrade the RNA.



samples 1B and 2A were split into 3 to become samples 1-3 and 4-6 respectively
sample 1A is sample 1B prior to removing genomic DNA with LiCl

Figure 7.4:

7.3.4 testing RNAse H on total RNA with DNA oligos complementary to the rRNA

Dec 14, 2007

Initially, I intended to use PCR product of 16S and 23S from genomic DNA using the primers on page 335. However, I tried a 100 μl reaction two times using the 16S *E. coli*, the 23S *E. coli*, and the 16S general primers, and I obtained little no DNA from these PCR reactions (reactions were for 35 cycles, annealing at 60C, extending 1 min at 72C):

Sample	primers	DNA (ng/ul)	260/280	260/230
1 (50 μl total volume)	16S general	6.0	2.14	1.13
2 (30 μl total volume)	16S *E. coli*	6.6	2.31	1.35
3 (30 μl total volume)	23S *E. coli*	24.5	1.90	1.91
4 (30 μl total volume)	16S general	0.5	0.34	0.07
5 (30 μl total volume)	16S *E. coli*	3.3	11.15	-0.75
6 (30 μl total volume)	23S *E. coli*	-0.9	0.65	0.18

I the table above wins the prize for the worst PCR spec values I've ever had.

My thesis defense was in 3 days, and I really wanted to try and give this RNAseH idea a stab before my talk (I'm filling this part of the notebook in early Jan after the defense chaos has subsided). As a last minute, trick I realized that the MICROBExpressB kit from Ambion actually has DNA oligos that are complementary to my rRNA sequence – that's how they pull the rRNA down for removal with the magnetic beads. So I decided to try and use the MICROBExpressB oligos with my rRNA. The only trick was that with this proprietary kit, I don't known *where* along the rRNA the sequences bind.

I grew two cultures in 4 ml of LB in a 12 ml falcon tube starting from a 1:50 dilution of overnite culture. I grew the samples 1 hr and 53 minutes to an OD600 of 0.523 and 0.537 respectively (not background subtracted). I prepared two total RNA samples using an RNeasy kit and I removed the genomic DNA using LiCl. The samples were resuspended into 30 μl of RNAse free H_2O . The yields were:

Sample	DNA (ng/ul)	260/280	260/230	yield
total RNA 1	2200.0	2.14	2.42	66 μg
total RNA 2	1932.7	2.14	2.44	58 μg

Using 10 μg of the above total RNA (the maximum amount recommended by the MICROBExpressB kit), I ran samples in RNAseH buffer (1a and 1b) and in MICROBExpress binding buffer (2a and 2b and 3). I heated the samples to 70C for 10 min (I used a 25 μl volume, so the samples were placed in a thermocycler). The RNAseH buffer samples were: 5 μl RNA (2 μg /μl), 4 μl MICROBExpressB oligo, 4.5 H_2O , 1.5 μl RNAseH buffer. For the MICROBExpress binding buffer: 5 μl RNA (2 μg /μl), 4 μl oligo, 6 μl MICROBExpress binding buffer.

I ran the samples on a 1% TBE gel. The first gel I ran with ssRNA sample loading buffer [NEB] (Figure 7.5A). The lanes ran completely wacko. I ran a second gel to see if I made the first gel wrong or something and I got the same poor migration. Before running the third gel, I did an EtOH precipitation to switch the buffers, because I had a hunch that the MICROBExpressB buffer was messing things up. The third gel (Figure 7.5B was fine.

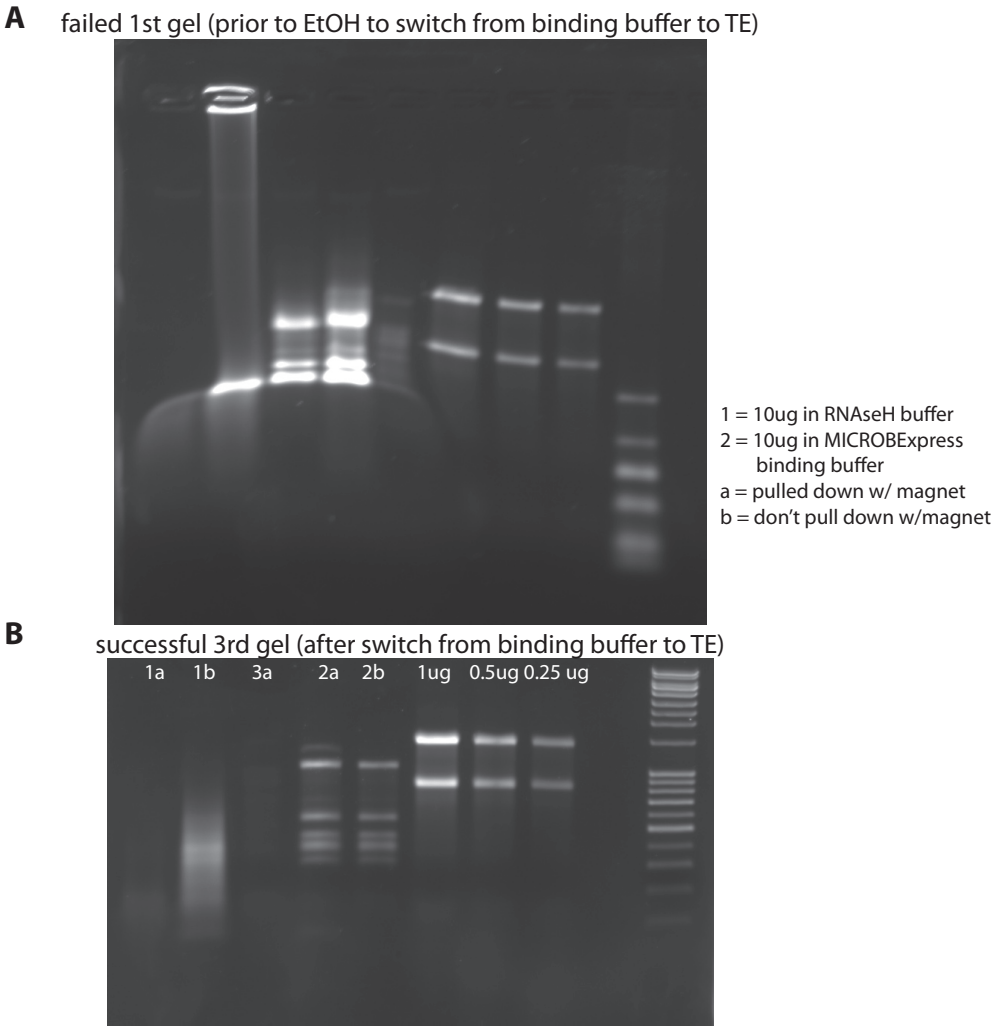


Figure 7.5:

Brief Conclusions: The RNAse H + MICROBExpress oligo seems to have work. The 23S and 16S bands are definitely cut in a systematic way (Figure 7.5. Looks like I need binding buffer like Ambion kit's MICROBExpress binding buffer (lanes 2a, 2b, and 3a), but I don't know what is in their buffer??? I guess they have more EDTA which prevents my sequence from getting degraded. Perhaps next run I'll try to anneal in EDTA, and then add RNAseH buffer+RNAse H. I should also run a few samples with the MICROBExpress binding buffer to test that too.

7.3.5 testing RNAse H on total RNA with DNA oligos designed by primer3

Mon Jan 7, 2008

Using the internal oligo design feature of Primer3, I designed 1000 oligos to the consensus sequence of 16S and 23S rRNA. The consensus sequence was created by aligning all of these rRNA sequences from *E. coli*and setting the nucleotide to N where there was not an exact match for the site across all rRNA species in *E. coli*. I then blasted the 1000 oligos for 16S and 23S against a database of all *E. coli*mRNA sequences using blastn. I then removed all oligos from the with an eval >25 (the purpose was to limit the nonspecific degradation that will occur if my DNA oligos binding RNA besides the rRNA I'm trying to destroy. I used a melting temp of 60C.

The primer3 files, blast database, and primer designs are available here. The final primers are contained in the files final_16S_primers and final_23S_primers.

As an initial test, I only ordered the primers that would allow my to chop the rRNA into thirds (i.e. the 23S will become 800bp and the 16S will become 500 bp if all primers cut successfully). If these tests are successful, I'll by all of the primers in a plate which should allow my to cut the rRNA into fragments of 80bp or less. I order the 23S primers at positions 931 and 1812; I ordered the 16S primers at positions 528 and 1078. For all four primers, I ordered the reverse complements as well so I can have a positive and negative control.

Wed Jan 9, 2008

I grew MG1655 in 2x4 ml of LB from a 1:50 dilution of overnite culture. After 2hr 38min, I took 2.5 ml samples and placed them in 5 ml of RNA protect. OD600 at this time was 0.697 and 0.739 for the two samples (not background subtracted). I used the RNeasy kit, eluted into 100 μl , ran a LiCl precipitation to remove the genomic DNA, resuspended into 31 μl of TE. The yields were:

Sample	DNA (ng/ul)	260/280	260/230	yield
sample 1	1921.9	2.18	2.28	59.6 μg
sample 2	1596.5	2.20	2.28	49.5 μg

how much primer to use?

I decided to use 1.5x of each oligo relative to the rRNA. I assumed that 16S and 23S each represent half of the total RNA population (e.g. in 10 μg of total RNA, 5 μg is 16S and 5 μg is 23S).

For the 23S rRNA, which is around 2500bp, I calculated:

$$20/2500 \times 5\mu g \times 1.5 \approx 60ng$$

At 100 μM the primers are at around 600 ng/μl ; at 10 μM they are around 60 ng/μl .

For this first experiment, I used 1 μl of 10 μM primer for both 23S and 16S. For samples 1-3, I used 10 μg of total RNA (from sample 1 above) in 25 μl of TES (TE + 50 mM NaCl). For samples

4-6, I used 10 μg of total RNA (from sample 2 above) in 25 μl of MICROBExpressB binding buffer (since this is what I knew worked last time). I heated the 6 samples to 70C for 10 minutes. To samples 1-3, I then added 25 μl of RNAse H buffer + 1.5 μl of RNAse H. To samples 4-6, I added 25 μl of MICROBExpressB binding buffer + 1.5 μl of RNAse H. I incubated all 6 samples at 37C for 15 minutes.

- Samples 1 and 4 contained appx 60 ng of each DNA oligo complementary to the 16S and 23S.
- Samples 2 and 5 contained appx 60 ng of the reverse complement of each DNA oligo (to check for not specific binding/degradation)
- Samples 3 and 6 contained no DNA oligo

After the RNAse H digestion, I cleaned up the reaction with EtOH precipitation, and resuspended the RNA pellets in 20 μl of TE. The yields for the six samples were:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**yield**
RNAse H 1	429.7	2.10	2.16	8.59 μg
RNAse H 2	418.2	2.10	2.15	8.36 μg
RNAse H 3	414.2	2.12	2.16	8.28 μg
RNAse H 4	456.7	2.10	2.14	9.13 μg
RNAse H 5	402.2	2.12	2.15	8.04 μg
RNAse H 6	399.9	2.14	2.18	8.00 μg

I ran 2 μl of each sample on

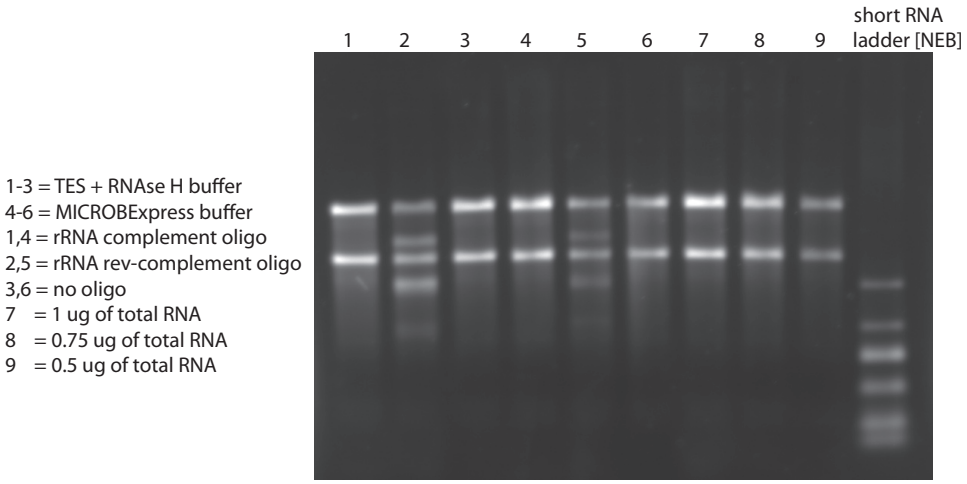


Figure 7.6: Not quite all the way degraded, but the RNAse H digestion certainly seemed to work.

Brief Conclusions: Looks like it worked! However, it looks like I had it backwards. The reverse complement oligos allowed it to cut, but the normal oligos did not. I'm very pleased with the specificity. True there were only 4 oligos in the mix, but there were a fair amount of them and they didn't seem to cut the RNA at all (compare lanes 1 and 3 and lanes 4 and 6, which contain the wrong oligo and no oligo respectively). On a more quantitative level, there is little to no difference between the RNA yields for the samples also (if there were no specific degradation, I'd expect the

yields to go down for the samples with DNA oligos). So far, this is better than I imagined. One problem is that the 16S and 23S bands are not digested to completion. I need to try different concentrations of oligo and RNAse H to figure out the optimal ratios for complete digestion.

7.3.6 optimizing RNAse H and DNA oligo concentration

Thur Jan 10, 2008

After my success with the custom-made oligos, I want to try and completely eliminate the original 23S and 16S rRNA bands. It's not clear if I need to add more oligo, more RNAse H, or if the gel in Figure 7.6 is as good as I'm gonna get.

sample	oligo amount (μl of 10 μM stock)	RNAseH (μl)
1	0.5	1
2	2	1
3	8	1
4	1	0.5
5	1	1.5
6	1	4.5

I ran a final sample (7) in which I only placed 1 μl of 16S oligo and 1 μl of RNAse H to make sure that the 23S rRNA remained uncut in the absense of 23S oligo. I used 5 μg of total RNA in each using the total RNA sample 2 from the section above.

The oligos and total RNA were placed in 25 μl of TES (50 mM NaCl + TE). I melted the RNA at 70C for 10 min. And then I added the 25 μl of RNAse H buffer plus the appropriate amount of RNAse H and incubated for 15 minutes at 37C. I cleaned up the reaction with EtOH and resuspended into 20 μl of TE. The yields for the 7 samples were:

Sample	DNA (ng/ul)	260/280	260/230	yield
RNAse H 1	244.0	2.12	2.12	4.88 μg
RNAse H 2	246.3	2.09	2.09	4.93 μg
RNAse H 3	329.4	2.03	2.03	6.59 μg
RNAse H 4	220.0	2.11	2.11	4.40 μg
RNAse H 5	220.7	2.11	2.03	4.41 μg
RNAse H 6	205.9	2.03	2.03	4.12 μg
RNAse H 7	211.5	1.97	1.97	4.23 μg

Note that as expected, the samples with the large amount of oligo added have a boost in concentration. I'm not sure why 4-7 have higher concentrations that 1-3.

I ran 2.5 μl of each sample on a 1% TBE gel with RNA sample buffer [NEB]. The gel was run at 120V for 50 min (Figure 7.7).

Brief Conclusions: Neither 23S or 16S was cut in this experiment (Figure 7.7). At first my confidence in this method was shattered. But then I had a look back at the experiment that worked (i.e. Figure 7.6 lanes 2 and 5) and I saw that it was the *reverse complement* oligos that worked. opps! At least I know that nonspecific degradation is still undetectable using different concentrations of RNAse H and the wrong oligos. I've already ordered an entire plate of appx 60 oligos tiled along the two rRNA genes, but now I'll have to order the reverse complements too. I can use the normal oligos as a good negative control. I'm also curious to know the specificity of this RNAse H digestion trick. Will it still degrade if my oligo has one mismatch? Does the location

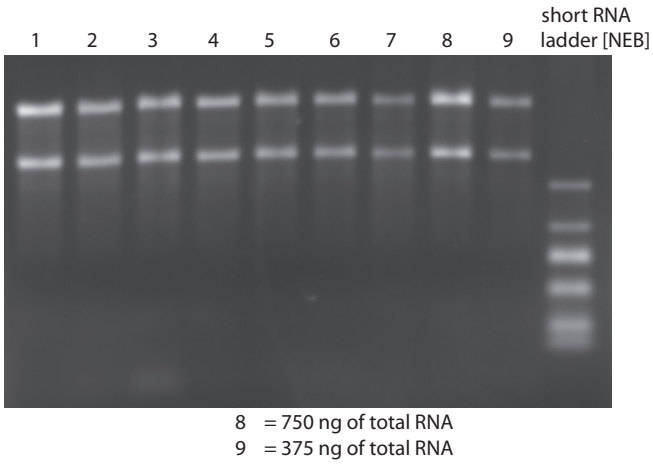


Figure 7.7: I used the wrong oligos. It was the reverse complement oligos that worked (Figure 7.6 lanes 2 and 5).

of the mismatch matter (presumably a mutation in the middle would prevent RNA binding and degradation while a mutation on one of the ends wouldn't change anything). Ok, now I need to redo this and figure out the proper concentrations of RNAse H and oligo...

7.3.7 optimizing RNAse H and DNA oligo concentration using reverse complement oligos

Fri Jan 11, 2008

The last section failed, because I used the forward oligos when I should've used the reverse complement ones. This experiment will be *exactly* the same 7 samples as above. However, I ran out of RNAse H, so that will come from a new tube and I ran out of total RNA, so I'll prep more and use the new sample.

I prepped two samples from overnite cultures using LiCl to remove the genomic DNA (see sections above for more info). Samples were resuspended into 30 μl of TE. I only spec'd the sample I used:

Sample	DNA (ng/ul)	260/280	260/230
sample A	1795.7	2.19	2.35

The remaining sample was placed at -20C and I'll spec it if I use it later.

Again the conditions for all 7 samples are the same as the previous section. The yields of the 20 μl of sample after EtOH precipitation were:

Sample	DNA (ng/ul)	260/280	260/230	yield
RNAse H 1	295.1	2.12	2.00	5.90 μg
RNAse H 2	322.8	2.04	1.80	6.46 μg
RNAse H 3	354.3	1.96	1.85	7.09 μg
RNAse H 4	254.3	2.10	1.96	5.09 μg
RNAse H 5	232.5	2.05	1.80	4.65 μg
RNAse H 6	264.4	2.07	1.90	5.29 μg
RNAse H 7	254.3	2.04	1.78	4.65 μg

2.5 μl samples were run on a 1% TBE agarose gel (Figure 7.8) for 50 minutes at 120 V.

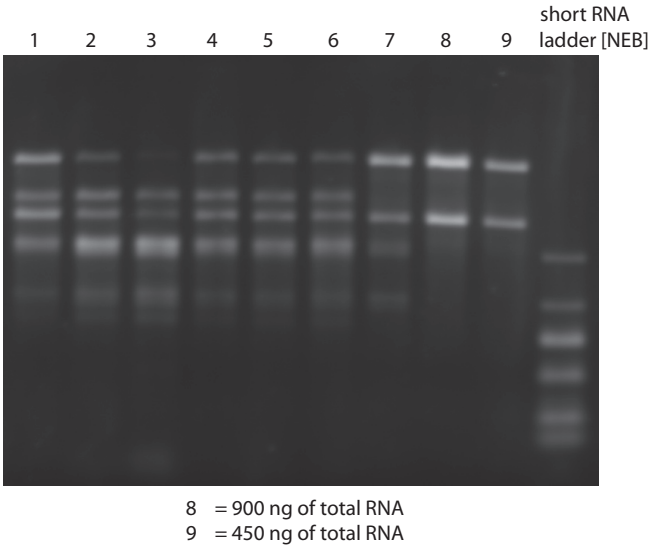


Figure 7.8: Different primer and RNAse H concentrations. Increasing primer concentration (lanes 1-3) seems more important than increasing RNAse H concentrations (lanes 4-6).

Brief Conclusions: Experiment worked much better when I placed the proper oligos in the tube. I attained practically complete degradation of the original 23S band and mostly complete degradation of the 16S rrrNA band (Figure 7.8 lane 3) when 8 μl of each 10 μM primer was added. So for practical purposes, I probably need to have my primers at higher concentration when I move to much higher numbers of primers. RNAse H seems like its good to go with 1 μl per rxn. Last, when I only placed the 16S oligo, I only degraded the 16S rRNA (lane 7) which is reassuring. Can't wait for the tiled oligo plate to arrive.

7.3.8 testing RNAse H on total RNA with a PLATE DNA oligos complementary to the rRNA

Thu Jan 17 13:29:10 EST 2008

Both the reverse complement and the forward rRNA oligo plates have arrived. Again info about the primers can be found here: The primer3 files, blast database, and primer designs are available here. The final primers are contained in the files final_16S_primers_w_rc and final_23S_primers_w_rc. Brief information about the scripts is in the file Notes.txt;

Today, I'll test the forward and reverse primer plates on the total RNA samples to see the effect on rRNA removal. There are 17 16S primers and 30 23S primers. The density of the primer tiling is such that it should chop the rRNA into fragments of 80bp or less.

I'm combining the primers into two separate tubes (one for 16S and one for 23S), so I don't have to individual add the large cocktail of primers one-at-a-time. I'm placing 5 μl of each 16S primer + 15 μl of TE for 100 μl of 10 μM 16S rRNA oligo stock. Unfortunately, the number of oligos for the 23S is so high that I can't create a 10 μM stock (I'd need to have the oligos delivered at a higher concentration, but this was the highest concentration they'd give me for a 25nM scale synthesis). For the 23S, I'm placing 5 μl of each for 150 μl of 6.6666 μM 23S rRNA oligo stock.

quantifying the total RNA

I made the genomic-DNA-free RNA in the previous section, but I never quantified it. I spec'd it with the nanodrop before I started the RNAse H experiment.

Sample	DNA (ng/ul)	260/280	260/230	yield
sample B	2485.7	2.15	2.30	74.6 μg

In the previous section, I found that adding more oligo help increase the amount of rRNA that was cut. I'm going to test the same thing in this experiment using all of the oligos form the rRNA oligo plates. Samples 1 and 2 will use the reverse complement plate, while 3 and 4 will use the forward plate. The two concentrations tested are the best performing from the previous section (i.e. 8 μl of 10 μM per rxn) and one-quarter of this optimal amount.

sample	plate used	oligo amount (μl of 10/6.66 μM stock)	RNAseH (μl)
1	revcomp	8,12	1
2	revcomp	2,3	1
3	forward	8,12	1
4	forward	2,3	1

The primers and total RNA were combined with NaCl to make a total volume of 25 μl TES. The TES mixture was heated to 70C for 10 minutes. A 1 μl RNAse H + 24 μl RNAse H buffer mixture was added to all four samples, and they were incubated at 37C for 15 minutes.

After the RNAse H digestion, 1 μl of glycoblue was added to each sample and they were cleaned up with EtOH, resuspended in 20 μl of TE, and spec'd with the nanodrop:

Sample	DNA (ng/ul)	260/280	260/230	yield
1	1222.1	1.77	1.99	24.4 μg
2	510.8	1.93	2.03	10.2 μg
3	1328.8	1.92	2.23	26.6 μg
4	507.2	1.99	2.13	10.1 μg

I prepared 2.5 μl of each sample to run on a 1% TBE gel (i.e. I mixed with RNA sample buffer [NEB]). I then added 1 μl DNAse-Turbo [Ambion] to samples 1 and 3, followed by a 10 min incubation at RT to see if I could get rid of some of the DNA oligo before running the samples on a gel. These were very unfavorable conditions for a DNAse reaction, since the sample was in TE, but we'll see. If it works in this situation, it should work in easier situations too (i.e. with less chelator).

I ran all 6 samples (4 normal + DNAse treated sample 1 and 3) on a 1% TBE gel for 1 hour (Figure 7.9).

Brief Conclusions: Clearly the DNA is making up the bulk of my reaction now as even after cleanup I have up to 5x more nucleic acid than starting RNA (see the four sample spec table above). However, it appears that the lower oligo amount performs as well as the high amount (compare lanes 1 and 2), so we can stick with this lower amount in the future. The rRNA bands are completely destroyed with the reverse complement oligos (Figure 7.9 lanes 1, 2, and 1*). However, the rRNA didn't get degraded all the way down to 80bp; there is a strong smear from 500bp-50bp. I think it the RNAseH rRNA removal technique is going to be the sole technique for rRNA removal, it'll need to be a two pass where I run the digestion, remove short RNA/DNA oligos, then run a second digestion with a second oligo plate (double the density, bringing it down to at best a 40bp window); with that tight a window, I'm actually going to be tiling 1/3 of the rRNA. Perhaps this is unnecessary if I do a two pass with a pull down (e.g. MICROBExpress) followed by a RNAseH digestion like this one. I need those Ambion RNA columns to come in; they're supposed to remove

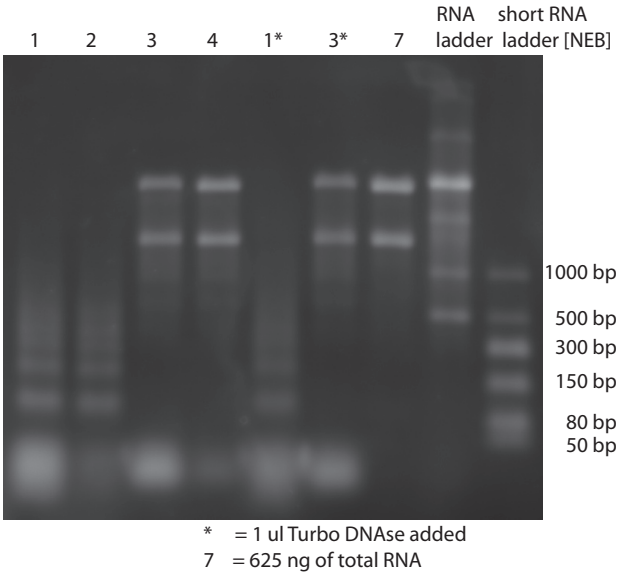


Figure 7.9: As expected the rRNA was degraded with the reverse complement oligo plate (lanes 1,2,1*).

small RNA pretty efficiently. I could also remove the short DNA downstream after the cDNA step. I do it anyways for the gel size-selection step. But I'd prefer to remove the short RNA earlier.

removing the short RNA with LiCl precipitation

Sat Jan 19, 2008

Rather than wait for the Ambion MEGACLEAR columns to arrive, I decided to precipitate the RNA from the above experiment with LiCl, which can also remove short RNA. Before starting a respec'd all of the samples to make sure the RNA hadn't degraded.

Sample	DNA (ng/ul)	260/280	260/230
1	1247.5	1.74	2.12
2	519.1	1.93	2.02
3	1244.6	1.88	2.28
4	520.2	2.04	2.06

The specs were quite similar to the previous specs, so looks like there's no problem with RNA degradation. I precipitated all four samples with LiCl (added 90 μl of H_2O and 50 μl of 7.5M LiCl) and eluted into 15 μl TE.

After the LiCl the specs were:

Sample	DNA (ng/ul)	260/280	260/230	yield (ng)
1	5.0	2.38	2.05	75
2	33.5	1.65	0.91	502.5
3	37.0	2.12	1.93	555
4	46.9	2.08	2.14	703.5

I ran the LiCl precipitated samples on a TBE agarose gel alongside some of the original sample for comparison (Figure 7.10).

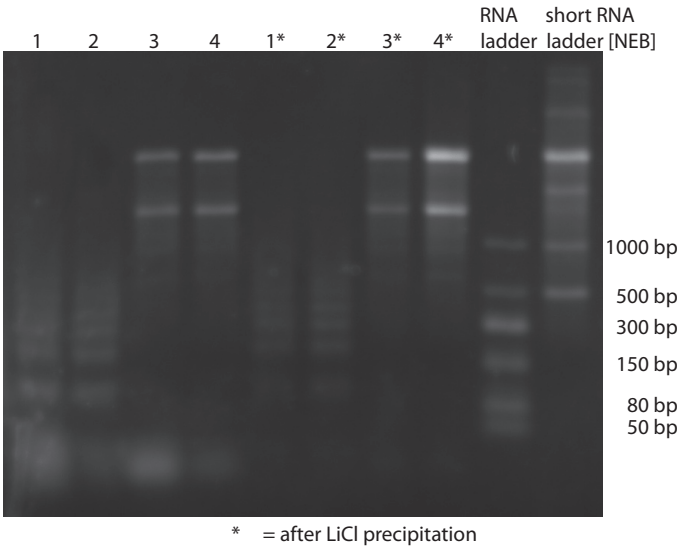


Figure 7.10: As expected the rRNA was degraded with the reverse complement oligo plate (lanes 1,2,1*).

Brief Conclusions: The LiCl certainly got rid of the DNA oligos (Figure 7.10 compare asterisk lanes vs non-asterisk lanes). It's hard to say if the precipitation also removed the short RNA fragment. It does appear like the RNA is less enriched at around 80bp after the LiCl (lanes 1* and 2*), but then again, the LiCl lanes are also quite a bit fainter, so it could be just do to the lane having less RNA overall. Hopefully the MEGAClear columns will do a better job and be easier.

7.3.9 running low on total RNA

I need to make more total RNA to keep pursuing these RNAse H based rRNA removal optimizations.

Mon Jan 21, 2008

I grew up 4 samples from a 1:50 dilution of overnite MG1655 culture. The cultures were grown in LB to an OD600 of 0.674, 0.661, 0.654, 0.586 respectively (OD not background subtracted). I added 2.5 ml of culture to 5 ml RNA protect, ran through a RNeasy RNA purification kit and used LiCl precipitation to remove genomic DNA.

I'll spec the samples as I use them for particular experiments downstream.

7.3.10 testing RNAse H on total RNA with a PLATE DNA oligos complementary to the rRNA and MEGAClear

Tue Jan 22, 2008

The Ambion MEGAClear columns have arrived, so hopefully they'll do the job of removing the low-MW RNA thats been chopped with RNAse H.

I'm running RNAse H samples, MicrobeExpress samples, and RNAse H + MicrobeExpress samples. The total RNA for samples 1-3 is from sample B on page 346. The total RNA from samples 4-7 is

from sample A in section 7.3.9 above. All MicrobeExpress samples used 50 μl beads and were run in PCR tubes in a thermocycler.

Sample	DNA (ng/ul)	260/280	260/230
sample A from section 7.3.9	2064	2.18	2.32

The set up is (RC = reverse complement):

1	5 μg total RNA; 2 μl 16S RC oligo mix; 3 μl 23S RC oligo mix; cleanup with EtOH; resuspend in 15 μl TE
2	5 μg total RNA; 2 μl 16S RC oligo mix; 3 μl 23S RC oligo mix; cleanup with MEGAClear; cleanup with EtOH; resuspend in 15 μl TE
2b	5 μg total RNA; 1 μl 16S RC oligo mix; 1.5 μl 23S RC oligo mix; cleanup with MEGAClear; cleanup with EtOH; resuspend in 15 μl TE
3	5 μg total RNA; 2 μl 16S forward oligo mix; 3 μl 23S forward oligo mix; cleanup with MEGAClear; cleanup with EtOH; resuspend in 15 μl TE (this shouldn't cut because it uses the forward primers)
4	run 10 μg total RNA through MicrobeExpress in 100 μl (50 μl binding + 50 μl beads); clean up directly in MEGAClear; EtOH; elute in 12 μl
5	run 10 μg total RNA through MicrobeExpress in 100 μl (50 μl binding + 50 μl beads); add 0.6 μl 16S RC oligo, 0.9 μl 23S RC oligo; heat to 70C for 10 min; bind at 37C and add 1 μl RNAse H; clean up directly in MEGAClear; EtOH; elute in 12 μl
6	run 10 μg total RNA through MicrobeExpress in 200 μl (150 μl binding + 50 μl beads); EtOH; elute in 12 μl
7	run 10 μg total RNA through MicrobeExpress in 200 μl (150 μl binding + 50 μl beads); add 0.6 μl 16S RC oligo, 0.9 μl 23S RC oligo; heat to 70C for 10 min; bind at 37C and add 1 μl RNAse HEtOH; elute in 12 μl

yields from the first four samples were:

Sample	DNA (ng/ul)	260/280	260/230	yield (μg)	loss
1	705.5	1.89	1.92	10.58	-112%
2	245.1	2.07	1.84	3.68	26%
2b	279.1	1.98	1.51	4.19	16%
3	305.9	1.99	1.54	4.59	8%

I ran 2.5 μl of each sample on a 1% TBE gel at 120V for 50 min (Figure 7.11).

yields from the second four samples were (estimated using Qubit values):

Sample	DNA (ng/ul)	260/280	260/230	Qubit (ng/μl)	yield (μg)	loss
4	72.3	1.81	1.20	67.7	0.81	92%
5	46.1	1.98	1.97	56	0.67	93%
6	174.1	2.07	1.90	120	1.44	86%
7	309.0	1.89	1.78	91	1.09	89%

I ran all of each sample on a 1% TBE gel at 120V for 50 min (Figure 7.12).

Brief Conclusions: The hybrid approach of using MICROBExpress followed by RNAse H looks like it might have real potential (Figure 7.12 lanes 5 and 7). The RNAse H looks like it helps remove more of the 23S and 16S bands. I'm not sure why the low-volume binding (lanes 4 and 5) doesn't have any of the high molecular weight RNA?

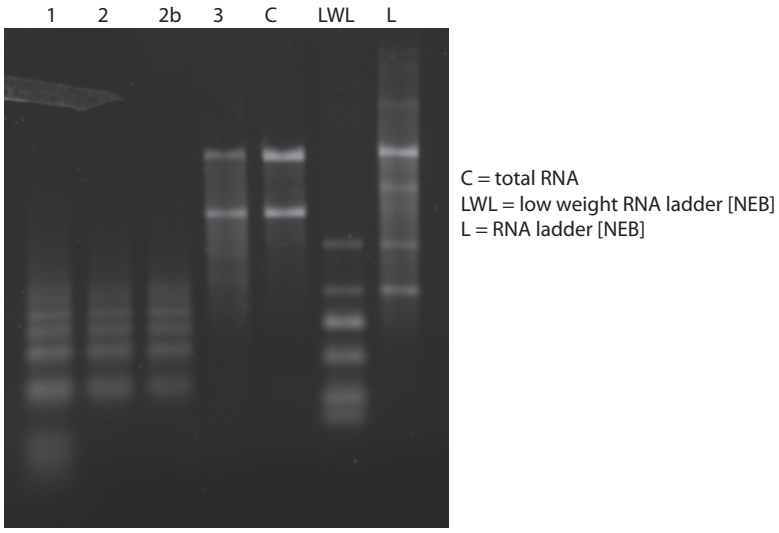


Figure 7.11:

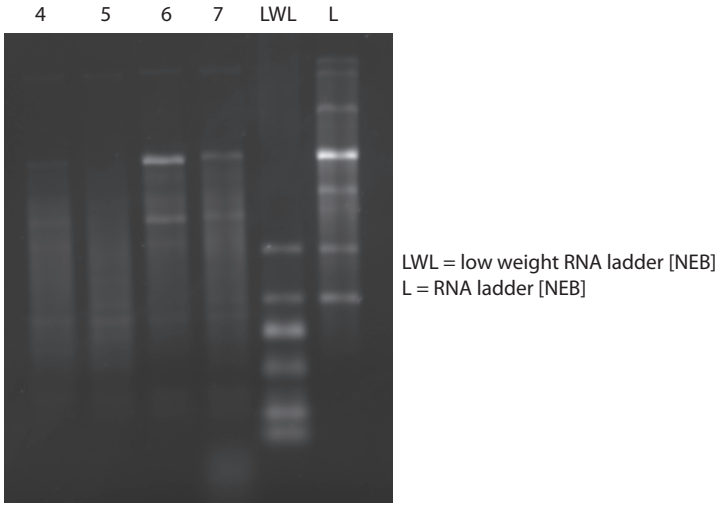


Figure 7.12:

7.3.11 cloning and sequencing to test the RNAse H based rRNA removal methods

The results in the previous section look promising enough that I want to try and clone some mRNA enriched cDNA to get some counts on the proportion of the remaining RNA that is still rRNA.

Wed Jan 23, 2008

All experiments with *oligos* used 1 μl 16S RC oligo mix and 1.5 μl of 23S oligo mix. All samples were resuspended into 12 μl of 0.5X TE. For the samples that required removal or loss of large amounts of RNA, I pooled multiple samples to have enough RNA to make cDNA and clone it. The number of samples pooled for each rRNA removal method is indicated in the table below.

The total RNA for this experiment was sample A from the section above and sample B (see spec below).

Sample	DNA (ng/ul)	260/280	260/230
sample B from section 7.3.9	2060	2.18	2.32

	description	pooled
1	10 μg total RNA; *oligos*; cleanup with MEGAClear; cleanup with EtOH	1
2	10 μg total RNA; 100 μl MICROBExpress (50 μl beads; 50 μl sample); *oligos*; cleanup with MEGAClear; cleanup with EtOH	2
3	10 μg total RNA; 200 μl MICROBExpress (50 μl ; 150 μl sample); *oligos*; cleanup with EtOH	2
4	10 μg total RNA; 200 μl MICROBExpress (50 μl beads; 150 μl sample); EtOH; *oligos*; cleanup with MEGAClear; cleanup with EtOH	3

The yields for the 4 samples (after pooling):

sample	Qubit RNA conc. (ng/μl)	yield (ng)
1	$>$ 200	-
2	97.8	1173
3	122	1464
4	61.4	736.8

I spec'd sample 1 with the nanodrop since it was too concentrated for the Qubit RNA dye:

Sample	DNA (ng/ul)	260/280	260/230	yield (ng)
1	660.6	2.08	1.82	7927

I made the cDNA as specified in the Preparation of PET libraries Protocol (section C.9 on page 427). I used all 11 μl for samples 2-5 with 1.5 μl superscript III; I used 5 μl of sample 1 with 3 μl superscript III. The following alterations were made to the PET cDNA library protocol: During 1st strand synthesis, I incubated at 50C for 45 minutes instead of 60. During 2nd strand synthesis, I incubated at 16C for 1hr 30 min instead of 2 hr. I did not heat inactivate the 2nd stand enzymes, rather I cleaned up the reaction directly with a Qiagen PCR purification column and eluted into 34 μl . I also did not heat inactivate the end-repair, as I also cleaned up this reaction directly with a Qiagen PCR purification kit and eluted into 30 μl . After which I quantified the DNA with the dsDNA HS Qubit kit:

sample	Qubit RNA conc. (ng/μl)	yield (ng)
1	35.2	1056
2	36.6	1098
3	89.6	2688
4	21.6	648

I used 1 μl (appx 2.1 μg) of BamHI adaptor.

Cloning the ds cDNA

Thur Jan 24, 2008

preparing the vector

I cut 2 μg of pUC19 with 1 μl BamHI, 2 μl 10x BSA, 13 μl H_2O , and 2 μl NEB3 buffer for 45 min at 37C. I cleaned up the digestion with a Qiagen PCR purification kit, eluted into 30 μl and spec'd the cut plasmid:

Sample	DNA (ng/ul)	260/280	260/230
cut puc19	51.3	1.86	2.18

I dephosphorylated 10 μl of the pUC19 with 7 μl H_2O , 2 μl antararctic phosphatase buffer, and 1 μl antaratic phosphatase at 37C for 30 min followed by heat inactivation at 65C for 5 minutes.

size-selecting the cDNA

I cleaned up the adaptored cDNA with a Qiagen PCR purification kit, eluted into 30 μl and ran all of it on a 65 ml TAE sybrsafe gel for 25 min at 90V with a PCR ladder and a 10-well wide comb. I gel purified with 550 μl of QG buffer and a Qiagen column.

I measured the final size-selected cDNA yields with a Qubit dsDNA HS kit:

sample	**Qubit DNA conc. (ng/μl)**	**yield (ng)**
1	3.65	110
2	2.73	81.9
3	10.0	300
4	5.26	158

Brief Conclusions: Once again, I was wishing I'd run the gel a little longer or a little hotter; Those adaptors are really in the way. Perhaps I should either use less adaptor or use the Purelink HS kit from invitrogen to try and eliminate more of the adaptor.

ligating the cDNA

I ligated the dephosphorylated vector to the adaptored cDNA using 10 μl of the gel purified cDNA, 2 μl of T4 ligase buffer, 2 μl of the dephosphorylated vector, 5 μl H_2O , and 1 μl of T4 DNA ligase at 16S for 30 minutes.

transforming the ligation

I transformed 2 μl of each of the four ligation into DH5alpha competent cells. I plated 50 μl and 100 μl of each on amp plates containing Xgal.

picking colonies

Fri Jan 25, 2008

I picked 24 colonies. I planned on picking 10 of each of the four samples, but I had no colonies for sample 3 and only three colonies for sample one.

second transformation

Since, I didn't have enough colonies, I retransformed an additional 2 μl of the ligation from the previous day. This time I used higher efficiency competent cells (oneshot TOP10 [Invitrogen]). I plated 100 μl and 200 μl of each of the four transformations. The TOP10 cells actually expired in Sep 07, hopefully they're still ok.

first minipreps

Sat Jan 26, 2008

I miniprepped all 24 samples.

second picking colonies

The higher efficiency TOP10 transformations worked well. Plenty of colonies for all four samples for the initial sequencing and for a potential downstream plate of sequencing. I picked 18 colonies (sample 1 and 3 since that was what was missing).

second minipreps

Sun Jan 27, 2008

I minipreped the 18 colonies picked yesterday.

spec/digest/sequence

I spec'd all 42 samples with the Nanodrop. The nomenclature: first number = sample ID (from one of the four rRNA removal techniques); second number = plate colony was picked from (e.g. 100μl is the plate with 100 μl from the DH5alpha transformation; T = TOP10); final letter = colony picked from the plate (in alphabetical order; colony 1 = a, colony 2 = b, etc. . .).

I digested 6 μl of 2 of each sample type with 1 μl of HindIII and 1 μl of EcoRI using EcoRI buffer just to verify that the inserts were of a decent size (Figure 7.13).

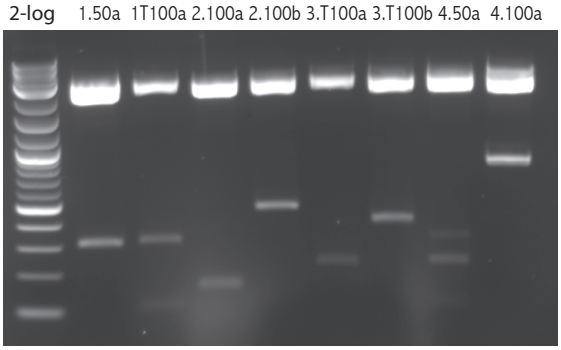


Figure 7.13:

The trimmed sequenced reads were blasted against the *E. coli*genome to determine the cDNA match. The raw sequence reads are available here and here. Note that I screwed up the nomenclature for the sample 4 when I submitted the sequence names to agencourt. I used 4-T100-a through 4-T100ul-j; The sample names in the table below are correct.

RNAse H rRNA removal results					
Sample ID	**ng/ul**	**260/280**	**260/230**	**yield (ug)**	**blast match**
1.50ul.a	286.39	1.95	2.26	14.3	rrlE 23S
1.50ul.b	233.78	1.93	2.2	11.7	rrlE 23S
1.50ul.c	214.58	1.97	2.23	10.7	rrsE 16S
1.T100ul.a	134.47	1.95	2.09	6.7	rrsE 16S
1.T100ul.b	218.73	1.97	2.25	10.9	rrlE 23S
1.T100ul.c	127.08	1.98	2.26	6.4	rrsA 16S
1.T100ul.d	171.97	1.97	2.26	8.6	rrsE 16S
1.T100ul.e	195.41	1.94	1.94	9.8	rrlA 23S
1.T100ul.f	179.58	1.99	2.21	9.0	rrsE 16S
1.T100ul.g	279.68	1.94	2.24	14.0	rrlE 23S
2.100ul.a	227.17	1.95	2.26	11.4	rrlE 23S
2.100ul.b	223.11	1.94	2.11	11.2	rrsE 16S
2.100ul.c	371.54	1.91	2.18	18.6	rrlA 23S
2.100ul.d	411.8	1.91	2.23	20.6	rrlA 23S
2.100ul.e	220.31	1.93	2.06	11.0	rrlE 23S
2.100ul.f	199.03	1.94	2.01	10.0	rrsE 16S
2.100ul.g	328.28	1.93	2.25	16.4	rrlE 23S
2.100ul.h	234.85	1.95	2.27	11.7	rrlE 23S
2.100ul.i	265.44	1.94	2.07	13.3	rrsE 16S
2.100ul.j	214.93	1.94	2.22	10.7	rrlE 23S
3.T100ul.a	152.71	1.94	1.88	7.6	rrlE 23S
3.T100ul.b	245.84	1.95	1.98	12.3	rrlE 23S
3.T100ul.c	158.01	1.98	2.2	7.9	rrlE 23S
3.T100ul.d	174.04	1.95	1.98	8.7	rrsE 16S
3.T100ul.e	87.93	1.95	2.05	4.4	-
3.T100ul.f	177.74	1.95	2.2	8.9	rrlE 23S
3.T100ul.g	159.02	1.95	2.17	8.0	rrlE 23S
3.T100ul.h	176.91	1.99	2.02	8.8	rrlE 23S
3.T100ul.i	116.5	1.97	2.13	5.8	rrlE 23S
3.T100ul.j	174.56	1.96	2.06	8.7	rrsE 16S
3.T100ul.k	257.99	1.94	2.22	12.9	-
4.50ul.a	261.6	1.94	2.25	13.1	rrlE 23S
4.50ul.b	386.9	1.91	2.22	19.3	rrlE 23S
4.50ul.c	257.63	1.94	2.25	12.9	rrlE 23S
4.50ul.d	306.41	1.93	2.19	15.3	rrlE 23S
4.50ul.e	328.23	1.93	2.23	16.4	rrlE 23S
4.100ul.a	261.63	1.95	2.25	13.1	mukF
4.100ul.b	274.39	1.95	2.22	13.7	mukF
4.100ul.c	218.56	1.97	2.28	10.9	rrlA 23S
4.100ul.d	315.31	1.89	1.81	15.8	cheA
4.100ul.e	280.3	1.94	2.27	14.0	rrlE
4.100ul.f	203.29	1.97	2.22	10.2	-

I thought it was weird that there were two mukF matches for sample 4. After an alignment (below), I think that I either picked the same colony twice or somehow two identical colonies were right next to each other on the plate (rather than two independent ligation events, which is what I'm really interested in).

```
CLUSTALW ALIGNMENT OF the two mukF cDNA matches

4-T100ul-g.trim      ----------GAGACACTTGCCGTGT-CAAAACCAGACAAGTGCCGCTGG 39
ecoli_genome         --------------------------CAAAACCAGACAAGTGCCGCTGG 23
4-T100ul-f.trim      ATCCGACCGAAGACAACTTGCCGTGTACAAAACCAGACAAGTGCCGCTGG 50
                                               **********************

4-T100ul-g.trim      ATCTTGGTCTGGTGGTACGCGAATATCTGTCACAGTATCCGCGTGCACGT 89
```

```
ecoli_genome          ATCTTGGTCTGGTGGTACGCGAATATCTGTCACAGTATCCGCGTGCACGT 73
4-T100ul-f.trim       ATCTTGGTCTGGTGGTACGCGAATATCTGTCACAGTATCCGCGTGCACGT 100
                      **************************************************

4-T100ul-g.trim       CACTTTGACGTTGCGCGTATTGTTATTGATCAGGCGGTACGTCTTGGCGT 139
ecoli_genome          CACTTTGACGTTGCGCGTATTGTTATTGATCAGGCGGTACGTCTTGGCGT 123
4-T100ul-f.trim       CACTTTGACGTTGCGCGTATTGTTATTGATCAGGCGGTACGTCTTGGCGT 150
                      **************************************************

4-T100ul-g.trim       AGCGCAAGCAGATTTCACCGGACTGCCAGCGAAATGGCAGCCGATTAATG 189
ecoli_genome          AGCGCAAGCAGATTTCACCGGACTGCCAGCGAAATGGCAGCCGATTAATG 173
4-T100ul-f.trim       AGCGCAAGCAGATTTCACCGGACTGCCAGCGAAATGGCAGCCGATTAATG 200
                      **************************************************

4-T100ul-g.trim       ATTACGGAGCCAAGGTACAGGCGCATGTCATCGACAAATATTGAACAAGT 239
ecoli_genome          ATTACGGAGCCAAGGTACAGGCGCATGTCATCGACAAATATTGAACAAGT 223
4-T100ul-f.trim       ATTACGGAGCCAAGGTACAGGCGCATGTCATCGACAAATATTGAACAAGT 250
                      **************************************************

4-T100ul-g.trim       GATGCCGGTTAAGCTGGCGCAGGCGCTGGCGAATCCGTTATTTCCGGCGC 289
ecoli_genome          GATGCCGGTTAAGCTGGCGCAGGCGCTGGCGAATCCGTTATTTCCGGCGC 273
4-T100ul-f.trim       GATGCCGGTTAAGCTGGCGCAGGCGCTGGCGAATCCGTTATTTCCGGCGC 300
                      **************************************************

4-T100ul-g.trim       TGGACAGCGCCTTACGTTCAGGACGCCATATTGGCCTCGACGAACTGGAT 339
ecoli_genome          TGGACAGCGCCTTACGTTCAGGACGCCATATTGGCCTCGACGAACTGGAT 323
4-T100ul-f.trim       TGGACAGCGCCTTACGTTCAGGACGCCATATTGGCCTCGACGAACTGGAT 350
                      **************************************************

4-T100ul-g.trim       AATCATGCATTCCTGATGGATTTTCAGGAATATCTGGAAGAGTTTTACGC 389
ecoli_genome          AATCATGCATTCCTGATGGATTTTCAGGAATATCTGGAAGAGTTTTACGC 373
4-T100ul-f.trim       AATCATGCATTCCTGATGGATTTTCAGGAATATCTGGAAGAGTTTTACGC 400
                      **************************************************

4-T100ul-g.trim       GCGTTATAACGTTGAGCTTATTCGCGCACCAGAAGGGTTCTTCTATTTAC 439
ecoli_genome          GCGTTATAACGTTGAGCTTATTCGCGCACCAGAAGGGTTCTTCTATTTAC 423
4-T100ul-f.trim       GCGTTATAACGTTGAGCTTATTCGCGCACCAGAAGGGTTCTTCTATTTAC 450
                      **************************************************

4-T100ul-g.trim       GCCCACGTTCCACCACGCTGATCCCTCGTTCCGTCTTGTCGGAACTGGAT 489
ecoli_genome          GCCCACGTTCCACCACGCTGATCCCTCGTTCCGTCTTGTCGGAACTGGAT 473
4-T100ul-f.trim       GCCCACGTTCCACCACGCTGATCCCTCGTTCCGTCTTGTCGGAACTGGAT 500
                      **************************************************

4-T100ul-g.trim       ATGATGGTCGGGAAAATCCTCTGTTATCTCTATCTCAGCCCGGAACGGCT 539
ecoli_genome          ATGATGGTCGGGAAAATCCTCTGTTATCTCTATCTCAGCCCGGAACGGCT 523
4-T100ul-f.trim       ATGATGGTCGGGAAAATCCTCTGTTATCTCTATCTCAGCCCGGAACGGCT 550
                      **************************************************

4-T100ul-g.trim       GGCGAATGAGGGGATTTTCACCCAGCAGGAACTGTACGACGAACTGCTCA 589
ecoli_genome          GGCGAATGAGGGGATTTTCACCCAGCAGGAACTGTACGACGAACTGCTCA 573
4-T100ul-f.trim       GGCGAATGAGGGGATTTTCACCCAGCAGGAACTGTACGACGAACTGCTCA 600
                      **************************************************

4-T100ul-g.trim       CCCTGGCCGATGAAGCAAAACTGCTGAAACTGGTGAACAACCGTTCAACC 639
ecoli_genome          CCCTGGCCGATGAAGCAAAACTGCTGAAACTGGTGAACAACCGTTCAACC 623
4-T100ul-f.trim       CCCTGGCCGATGAAGCAAAACTGCTGAAACTGGTGAACAACCGTTCAACC 650
                      **************************************************
```

```
4-T100ul-g.trim      GGTTCAGACGTTGACCGTCAGAAGTTGCAGGAGAAAGTACGTTCTTCGCT 689
ecoli_genome         GGTTCAGACGTTGACCGTCAGAAGTTGCAGGAGAAAGTACGTTCTTCGCT 673
4-T100ul-f.trim      GGTTCAGACGTTGACCGTCAGAAGTTGCAGGAGAAAGTACGTTCTTCGCT 700
                     **************************************************

4-T100ul-g.trim      CAACCGTCTGCGTCGTTTAGGCATGGTGTGGTTGTCGGATTCGGATCCTC 739
ecoli_genome         CAACCGTCTGCGTCGTTTAGGCATGGTGTGGTT----------------- 706
4-T100ul-f.trim      CAACCGTCTGCGTCGTTTAGGCATGGTGTGGTTGTCGGATTCGGATCCTC 750
                     *********************************

4-T100ul-g.trim      TAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTG 789
ecoli_genome         -------------------------------------------------
4-T100ul-f.trim      TAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTG 800


4-T100ul-g.trim      TTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGC 839
ecoli_genome         -------------------------------------------------
4-T100ul-f.trim      TTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGC 850


4-T100ul-g.trim      CGGAAGCATAAAGTGTAAAGCCTGNGGTGCCTAATGAGTGAGCTAACTCA 889
ecoli_genome         -------------------------------------------------
4-T100ul-f.trim      CGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCA 900


4-T100ul-g.trim      CATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCG 939
ecoli_genome         -------------------------------------------------
4-T100
```

Brief Conclusions: That really sucked!

I was hoping to have a range of success rates with each of the four rRNA removal strategies. But three of them failed to produce a single mRNA in 10 sequencing reactions (see table above). Method 4, using 200 μl MICROBexpress, EtOH; RNAse H + *oligos*; cleanup with MEGAClear; and cleanup with EtOH performed the best with 2 in 9 samples being mRNA (22%). I'm going to sequence a half a plate or more of these samples to try and determine more precisely the true rRNA proportion. Since I didn't run any samples without the RNAse H, its hard to know if the RNAseH really helped or not (it appeared to do so in the gel in the previous section, but

Thoughts for next round:

try: 1) two rounds of MICROBExpress with EtOH (as I did before to get around 28%); 2) two round of MICROBExpress with EtOH followed by RNAseH and MEGAClear (i.e. is the RNAse H helping at all?); 3) one round of MICROBExpress followed by 2 rounds of RNAse H (using the two different RC plates and a MEGACLEAR in between); 4) three rounds of MICROBExpress with EtOH

In addition, I'm working on biotinylating the oligos from the oligo plate to see how well it works to try MICROBExpress followed by a second type of pull down with a different oligo set (or perhaps 3 different oligo sets).

All of this will require quite a large number of sample poolings to ensure I have enough cDNA to clone.

sequencing more of technique 4

Tue Feb 12, 2008

I want to see how well these crude statistics estimated on the 10 quicklane samples extrapolate to a larger sample size; in addition, I need to test if my cells/plates are compatible with the 96-well single-pass sequencing at agencourt. Towards this, I picked 96 colonies of the sample 4, which showed 22% mRNA. The preparation process for agencourt is beautiful; just pick 96 colonies into a 96-well skirted PCR plate filled with LB with 10% glycerol, grow statically at 37C for exactly 12 hours and then freeze at -80C and ship to agencourt. Dead easy, and very nice to not have to do 96 minipreps!

Wed Feb 19, 2008

Although it was easy, it wasn't particularly rapid (though they explicitly state that it should take around a week and it did) to get sequences done this way. On the whole all but 2 wells were successful sequenced, which is fantastic. In mapping the reads back, it became quite clear that the low sampling I used with the 10 quicklane reads was a statistical anomaly. The read mRNA rate based on this plate was 4 in 94 (4.26% mRNA).

The sequence data is here.

This excel file contains the BLAST results summary.

7.3.12 specificity of RNAse H oligos

I want to know how much non-specific degradation I'm getting from the oligos. I purposely chose oligos that had the least similarity with genomic regions, but I still don't know how well an oligo has to match to allow the RNAse H to cut. To gain a little insight into this potential degradation, I developed 10 versions of the center-most 16S oligo from the first reverse complement oligo plate. Successful directed RNAse H degradation at the oligo should roughly break the 16S in half which makes for an easy gel based assay of degradation ability. Each variant has a mutation at one location along the primer. The changed nucleotide position is labeled with a * and the randomized order number for the primer is presented on the left:

```
# original
4) TTTACGGCGTGGACTACCAG

# two center changes
        # change G to anything else (one at a time)
2)      TTTACGGCGTCGACTACCAG
6)      TTTACGGCGTAGACTACCAG
9)      TTTACGGCGTTGACTACCAG
                 *

        # change T to anything else (all together)
1)      TTTACGGCGVGGACTACCAG
                *

# 5' changes
        # 5 from the end
```

```
7)      TTTADGGCGTGGACTACCAG
            *

        # 3 from the end
10)     TTVACGGCGTGGACTACCAG
          *

        # 1 from the end
3)      VTTACGGCGTGGACTACCAG
        *

# 3' changes
        # 6 from the end
11)     TTTACGGCGTGGACVACCAG
                      *

        # 3 from the end
5)      TTTACGGCGTGGACTACDAG
                        *

        # 1 from the end
8)      TTTACGGCGTGGACTACCAH
                         *
```

1 center T→V
2 center G→C
3 5' 1bp
4 original
5 3' 3bp
6 center G→A
7 5' 5bp
8 3' 1bp
9 center G→T
10 5' 3bp
11 3' 6bp

should I try shorter lower MT oligos?

7.3.13 initial testing of the 16S mutation oligos

Thur Feb 7, 2008

I tested all 11 variants (including the non-mutation version) using the standard RNAse H protocol I've been following: 4 μl of 10 μM oligo, 5 μg of total RNA, in 25 μl total volume TES; 10 min 70 C; add 25 μl RNAse H buffer, H_2O , with 1 μl RNAse H; 37 C for 15 minutes.

For the total RNA, I spec'd and used sample C from section 7.3.9:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**
sample C from section 7.3.9	2982.5	2.15	2.28

I cleaned up all samples with EtOH and glycoblue and eluted into 15 μl of TE:

Sample	DNA (ng/ul)	260/280	260/230
1	282.3	2.04	1.82
2	290.6	1.98	1.81
3	256.3	2.07	2.14
4	271.1	1.89	2.05
5	332.7	1.85	1.46
6	258.0	2.16	2.34
7	270.1	2.07	1.95
8	269.5	2.12	2.10
9	265.7	2.11	2.10
10	264.6	2.09	2.05
11	287.1	2.02	1.63

I ran 2.2 μl of all 11 samples on a 1% TBE agarose gel for 50 min at 120V (Figure 7.14).

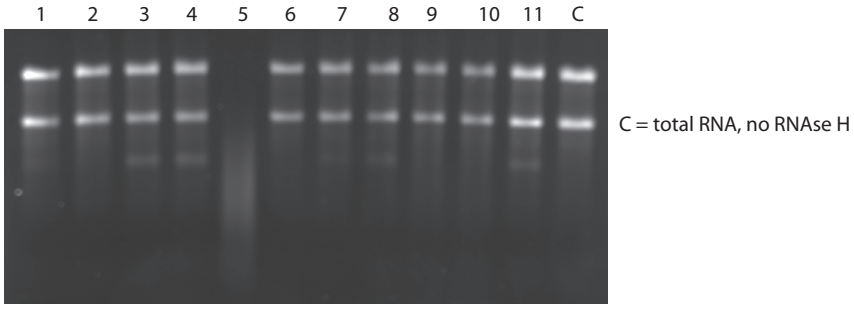


Figure 7.14: mutation 16S oligos with RNAse H; what's going on with the oligo in lane 5 (3' 3bp)? The lanes are: 1) center T→V; 2) center G→C; 3) 5' 1bp; 4) original; 5) 3' 3bp; 6) center G→A; 7) 5' 5bp; 8) 3' 1bp; 9) center G→T; 10) 5' 3bp; 11) 3' 6bp

Brief Conclusions: The original unmutated primer cut as expected (Figure 7.14 lane 4) – although it didn't cut as completely as I expected. If the mutation is in the middle (i.e. position 10 or 11 of the 20-mer oligo) the oligo doesn't noticably cut the 16S band regardless of the substitution (Figure 7.14 lanes 1, 2, 6, 9). On the other extreme, a subtitution on the last basepair on either end of the oligo (i.e. position 1 or 20) results in cutting (lanes 3 and 8). The remaining primers all seem to have cut to some extent except the 5' 3bp mutation (lane 10), which is weird because the 5' 5bp mutation (which should be less able to bind than the 3bp version) did cut.

The complete odd ball is lane 5 where the 3' 3bp completely sheared the RNA. Did some RNAse get in there? Is the oligo perhaps poorly synthesized so that it contains a large number of smaller oligos that bind everywhere (and result in non-specific RNA degradation?). Is this type of thing common with any large set of synthesized oligos, so that I need to individually screen my oligo library for the "good" ones?

further testing of the oligo from the degraded RNAse H lane

Mon Feb 11, 2008

I want to see if the RNAse H primer have potentially nasty oligos that for some reason lead to the general degradation seen in lane 5 in Figure 7.14. I plan use the mutation primers 4 (original) and 5 (3' 3bp). I'll clean each up with a nucleotide removal kit [Qiagen] to remove any RNAses or short nucleotide fragments that might be leading to general degradation, and I'll run an RNAse H rxn the clean and the original samples.

sample	description	amount to use for RNAse H
1	3' 3bp	240 ng
2	original mutation primer	240 ng
3	16S plate primer	1000 ng
4	23S plate primer	1700 ng

I cleaned up and eluted each sample into 30 μl of EB buffer; the specs for the cleaned (a) and uncleaned (b) samples were:

sample	**conc (ng/μl)**	**260/280**	**260/230**
1a	135.2	1.80	2.18
1b	288.4	1.80	1.21
2a	135.7	1.84	2.33
2b	318.1	1.82	1.33
3a	323.0	1.72	1.98
3b	890.3	1.76	1.40
4a	355.0	1.76	1.89
4b	1452.4	1.71	1.23

I ran the standard RNAse H reaction (TES 10 min 70C; add RNAse H buffer 1 μl RNAse H, 15 min 37C). Cleaned up with EtOH and resuspended in 15 μl of TE. Unfortunately, the first RNAse H reaction, I got the my 5M ammonium acetate and the TE elution buffer mixed up and ran the RNAse H reaction in acetate. More unfortunately, I also realized when I noticed this mistake that I'd also used the acetate instead of the TE for the resuspension of my total RNA after LiCl, so those are probably ruined. The acetate and the TE have exactly the same ambion tube style, so I got them confused (now the acetate is clearly marked to prevent another mistake).

The specs from the acetate mistake reactions:

sample	**conc (ng/μl)**	**260/280**	**260/230**
1a	271.7	1.88	1.82
1b	317.2	1.86	1.79
2a	194.8	1.81	1.71
2b	283.3	1.87	1.84
3a	251.0	1.84	1.76
3b	357.6	1.84	1.83
4a	426.7	1.83	1.81
4b	412.6	1.83	1.84

After I reran the RNAse H reactions using the correct TE buffer rather than RNAse H the specs were:

sample	conc (ng/μl)	260/280	260/230
1a	307.7	2.03	1.85
1b	335.3	2.06	1.85
2a	325.6	2.07	1.89
2b	326.7	2.06	1.88
3a	350.9	2.05	2.00
3b	357.7	2.02	1.91
4a	388.9	2.01	1.90
4b	407.2	1.99	1.90

I ran 2.5 μl of all 8 samples for the acetate mistake and the no-mistake RNAse H mutation test on a 1% TBE gel for 45 min at 120V (Figure 7.15.

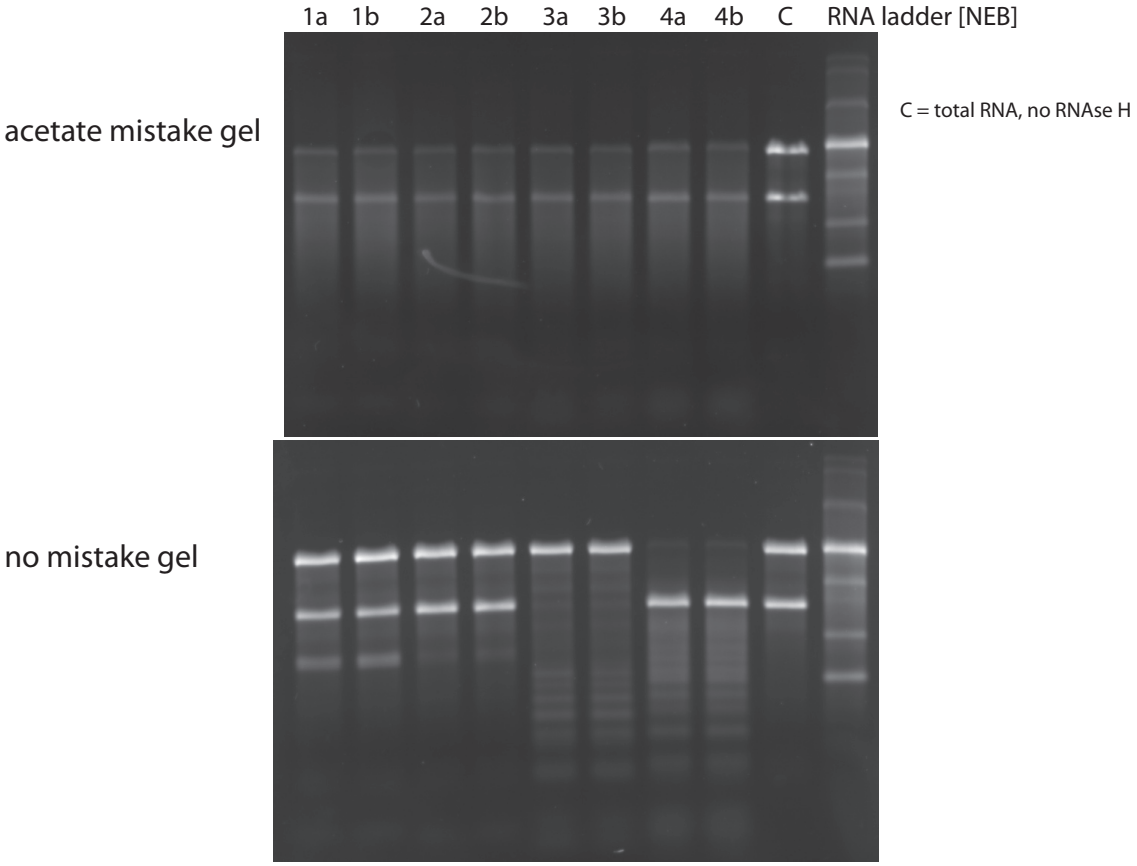


Figure 7.15:

Brief Conclusions: The RNAse H didn't function in the acetate mistake (of course RNAse H was inhibited Figure 7.15). The 16S alone and 23S alone are beautifully clean and have somewhat reinvorated my hope that this RNAse H technique might just work. On the 16S only lanes (3a and 3b), The 23S band looks quite clean and undegraded. Similarly for the 23S primers in lanes 4a and 4b. However, the primary point of this follow up experiment was to see what was going on with the sample in lane 5 in Figure 7.14. In this case neither of the two samples using the 3' 3bp primers from that sample caused a similar degradation. Did some RNAses get in that sample last time?

7.3.14 cloning and sequencing to test additions RNAse H and oligo pulldown combinations

Given my marginal success in the previous sequencing run with the RNAse H + 1x MICROBExpress, I'm going to try variations on that theme (again the best performing was: 10 μg total RNA; 200 μl MICROBExpress (50 μl beads; 150 μl samples; EtOH; *oligos*; cleanup with MEGAClear; clean up with EtOH; pool 3 samples to have enough cDNA). The RNAse H based success had an rRNA:mRNA ratio of 5:1 (i.e. 20% mRNA) that I previously attained using two rounds through the MICROBExpress kit, so now I want to try out multiple MICROBExpress runs combined with the RNAseH. In addition, I have a second 16S and 23S oligo tiling plate that should allow even further degradation of the rRNA, so I want to toss that in as well to give it a shot.

	description	**pooled**
1	10 μg total RNA; 2x 200 μl MICROBExpress (50 μl beads; 150 μl sample); EtOH;	2
2	10 μg total RNA; 2x 200 μl MICROBExpress (50 μl beads; 150 μl sample); EtOH; *oligos*; cleanup with MEGAClear; cleanup with EtOH	3
3	10 μg total RNA; 2x 200 μl MICROBExpress (50 μl beads; 150 μl sample); EtOH; *oligos*; *oligos plate 2*; cleanup with MEGAClear; cleanup with EtOH	4

Method one is the previous best. Method two is similar to the method four in my last RNAse H sequencing tests but with the addition of a second MICROBExpress rRNA removal step. If method one and two perform similarly, then perhaps the RNAse H isn't helping anything. Method three adds the second RNAse H tiling oligo plate to try and further reduces the rRNA size.

preping more total RNA

Sun Feb 10, 2008

I grew for MG1655 cultures in LB for 2 hr 30 min from a 1:50 dilution of overnite culture to OD600 of: 0.671, 0.672, 0.723, 0.722.

These samples will be total RNA samples e, f, g, h (always restarting with a,b,c is making my freezer too confusing.

removing the rRNA

Tue Feb 12, 2008

I planned on using samples e and f to try the three rRNA removal techniques in the above table. However, when I went to spec the RNA before I started I realized that the TE/acetate confusion that happened with the mutation primers had also sneaked into my total RNA. That is, instead of eluting the total RNA into 30 μl of TE, I eluted it into the high concentration sodium acetate that is used for EtOH precipitation (the tubes looked exactly the same). This experiment will have to be postponed.

preping even more total RNA

Wed Feb 13, 2008

I grew for MG1655 cultures in LB for 2 hr 30 min from a 1:50 dilution of overnite culture to OD600 of: 0.718, 0.674, 0.725, 0.722, 0.708, and 0.787. These represent total RNA samples i, j, k, l, m, and n.

This time I wanted to be sure that everything would go smoothly with the rRNA removal tomorrow, so I spec'd samples i and j:

Sample	RNA (ng/ul)	260/280	260/230
sample i	2923.9	2.11	2.21
sample j	1765.1	2.19	2.34

rRNA removal and cDNA generation

Thu Feb 14, 2008

I ran the 3 different rRNA removal strategies from the table above using 10 μg total RNA samples i or j (3.4 μl of i or 5.6 μl of j). As indicated in the table, I pooled 2, 3, and 4 replicates of samples 1, 2, and 3 respectively due to the large potential for RNA loss with each rRNA removal. I used 0.5 μl of the 16S oligo and 0.75 μl of the 23S oligo for each oligo removal step. The MICROBExpress reactions were not cleaned up in between the two replicate removals. Rather I just added more oligo and started again. With the first MICROBExpress removal, I used 4 μl oligo, 150 μl of binding buffer, and 50 μl of beads (very similar to the actual Ambion protocol, but with 50 μl less binding buffer so it fits nicely in a PCR tube). The second reaction, after the large majority of the RNA had been removed in the first round, used only 2 μl of oligo and 25 μl of beads. The elution buffer for the MEGAClear was 0.1 mM EDTA. So in between the two RC oligo/RNAse H steps for sample 3, I added enough EDTA to bring it up to the normal TE concentration of 1 mM. After all of the pooling and cleanups, I quantified the RNA with the Qubit:

Sample	RNA (ng/ul)	yield
rRNA removal strategy 1	102.4	1433 ng
rRNA removal strategy 2	7.8	109.2 ng
rRNA removal strategy 3	16.7	233.8 ng

I used 12 μl of all of the samples with 1.5 μl of Superscript III for 1st strand synthesis. The 1st strand synthesis was run for 1 hour at 50C.

The yields after 2nd strand synthesis and end-repair were:

Sample	yield	amount of adaptor
rRNA removal strategy 1	570.6 ng	0.3 μl (600 ng)
rRNA removal strategy 2	270 ng	0.1 μl (200 ng)
rRNA removal strategy 3	77.4 ng	0.045 μl (90 ng)

Pretty low amounts of cDNA to work with, but not too surprising given the low RNA amounts to start with. Given lower DNA amounts and the difficult of gel selection when there is a ton of adaptor, I used less adaptor (somewhere between 1:50 and 1:100).

cDNA size-selection, vector prep, and cloning

Fri Feb 15, 2008

I decided to run the gel a little longer to try and separate the adaptors from the cDNA a little more. Since this will require me to cut a longer stretch of gel, I tried to lessen the amount of agarose

in a few ways: 1) I used a 0.8% TAE agarose gel with SYBR safe (usually I use 1%); 2) I used a minelute column for the final cDNA cleanup[1], with less volume of sample, I used a thinner 45 ml gel (usually I need a 60 ml gel). I used a 12-well comb with 1.5 mm wells to make the column slimmer (usually I use a 10-well with 1.5 mm wells). 3) I ran the gel for 25 min at 100V (previously I used 90V).

Something in that mix really helped (maybe the minelute? or the lower adaptor amount?), because the adaptors were not visible on the gel AT ALL. I could clearly see the smear for all of the cDNA sample (though sample 3 was faint), and I cut from 300bp to around 3000bp for all of them. Next time I think I can use an even thinner gel (35 or 40 ml) for even less agarose.

I cleaned up the gel size-selected cDNA with a Qiagen gel column cleanup, eluted into 30 μl , and quantified the cDNA with a Qubit dsDNA HS assay:

Sample	**DNA (ng/μl)**	**yield (ng)**
gel cleanup from removal strategy 1	7.6	228
gel cleanup from removal strategy 2	4.16	124.8
gel cleanup from removal strategy 3	1.004	30.12

The gel selection loss was pretty typical at around 50%.

Preparing the pUC19 vector

The vector was prepared as in the previous cDNA cloning. The pUC19 vector was cut: 2 μl pUC19 (2 μg), 1 μl BamHI, 2 μl 10X BSA, 13 μl H_2O , 2 μl NEB3 for 45 min at 37C. The digestion was immediately cleaned up with a Qiagen PCR purification kit and eluted into 30 μl . The spec was:

Sample	**RNA (ng/ul)**	**260/280**	**260/230**
pUC19	42.4	1.88	2.17

I took 10 μl of this BamHI cut vector with 7 μl H_2O , 2 μl antarctic phosphatase buffer, and 1 μl antaractic phosphatase to dephophorylate the vector for 30 min at 37C followed by deactivation of the phosphatase at 65C for 5 minutes.

Ligating and cloning the cDNA

The ligation for all 3 rRNA removal strategies was: 2 μl T4 ligase buffer, 2 μl dephosphorylated pUC19 vector, 10 μl cDNA, 5 μl H_2O . 30 min at 16C; I transformed 2 μl of each ligation into One-Shot TOP10 cell [Invitrogen]. I plated 75 μl and 150 μl of each onto amp plates with x-gal for blue-white cloning.

Sat Feb 16, 2008

Plenty of colonies: life is good.

I sequenced one-third of a plate for all three strategies. Strategies 1-3 yield mRNA concentrations of 23% 10% and 23% respectively (see this excel file with the BLAST results summary for details).

Brief Conclusions: Appears that the RNAseH isn't really helping much. Nice to see I can still attain 20% mRNA or greater when I sequence larger numbers of candidates. I think this suggests I should focus on pulldown methods and give up on the beautiful but ineffective idea of RNAseH.

[1] I tested the sample loss with the minElute -vs- the standard cleanup kit from Qiagen (see section 10.8 on page 404). The minElute loss is around 20% and the standard cleanup loss is around 10%. However, I'd guess that the extra loss from the minElute is more than made up for by the higher yield that will be obtained when the gel cleanup has less agarose.

7.4 rRNA removal via oligo pulldown

The Ambion rRNA removal kit works with a oligo that is captured in a sandwich complex with another biotinylated piece of DNA. I don't think they target many sites along the 16S and 23S rRNAs, so can I do better if I use the rRNA oligos designed for the RNAseH trick. I can biotinylate the oligo mix, capture the rRNA and pull it down similar to the MICROBExpress kit, but with wider coverage and no requirement for a sandwich complex to capture the oligo/rRNA complex.

7.4.1 biotinylating the oligo mix

I'm going to try and biotinylate the 23S oligo mix from rRNA RC plate 1 and the biotinylation labeling kit that comes with the Affymetrix arrays (they claim 90% labeling efficiency). I tried to label 2.8 μg of the oligo with 10 μl of terminal transferase buffer (TdT), 2 μl of the Affymetrix labeling reagent (presumably biotinylated oligo nucleotides), and 2 μl of TdT. I cleaned up the reaction with a nucleotide cleanup kit [Qiagen] and eluted into 30 μl of EB. I spec'd the two samples:

Sample	DNA (ng/ul)	260/280	260/230
23Sa	1.66	1.41	0.82
23Sb	1.54	1.89	0.95

Despite this incredibly poor yield (where'd my DNA go?), I ran 15 μl of each sample without neutravidin and 15 μl of each sample with neutravidin; it there was sufficient DNA and it was labeled I should see a gel shift with the neutravidin bound sample. I ran a 3.5% TBE Nusieve gel for 40 min at 110V (was too long) (Figure 7.16).

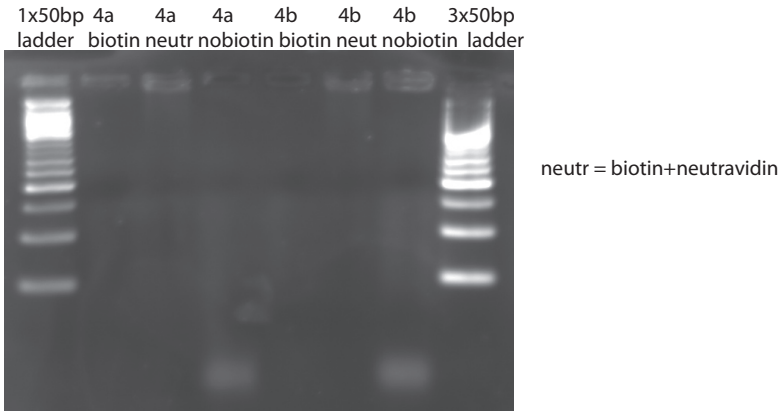


Figure 7.16:

Brief Conclusions: Well you can see the primer in the no biotin lane, which suggests I should run the gel for less time and use a 25bp and/or 10bp ladder next time. If you really use your imagination, there might be some gel shifted DNA towards the top of the two neutravidin lanes (4a neutr and 4b neutr)? The main problem of course is that I lost too much of my oligo somehow. I'd guess it was during the Qiagen cleanup. I did the cleanup to remove the unincorporated biotinylated nucleotides from the sample, so that if I eventually used them with beads, the nucleotides wouldn't bind up all of the streptavidin. Perhaps I can use EtOH instead to hopefully have less loss and allow higher concentrations of biotinylated oligo to be processed.

7.5 rRNA removal via ultracentrifugation

This idea was abandoned after I couldn't develop a total RNA extraction protocol that wouldn't use harsh chemicals that would denature the ribosome. I could either make a gentle protocol that resulted in completely degraded RNA or I could make a harsh chemical protocol that resulted in degraded ribosomes.

7.6 rRNA removal via column chromatography

This idea was abandoned for the same reason that ultracentrifugation was abandoned.

Chapter 8

USER cloning

THIS CHAPTER/PROJECT IS NOT ACTIVE

I still think the idea is useful; but it was taking up time that I didn't have. I need to graduate someday.

Hopefully this will be a short successful chapter. All I want to do is get a cloning proceedure that works without restriction enzymes. The TA-TOPO kits do this, but they aren't directional. The USER kit from NEB kinda does this, but it isn't general. It requires a 9-mer overhang thing that while cool, is a little long for what I want to do. Their 9-mer eliminates the need for ligase. The method I want to get going here will use ligase.

8.1 Motivation

I think the original motivation for USER cloning was that it was faster and didn't require the ligation step. I am more interested in a restriction enzyme free ligation procedure. There are a few situations where it would be nice to not have to cut your DNA with a restriction enzyme for example:

1. you want to clone a gene, but one (or both) of the enzymes at for your insert sites have a site inside your insert. you either have to switch the insert sites on the vector or somehow mutate the site in your gene and hope it doesn't mess up the way the gene works.
2. you are cloning random or randomized DNA where you can possibly know if you have the sites
3. you are cloning large pieces of DNA where you almost certainly have one of the sites
4. you want to circularize a big piece of DNA without cutting it or ligating on adaptors

I've encountered many of those problems already and the others, I see myself bumping into in the future as I try fancier ways to do DNA sequencing.

The basic idea behind USER cloning is shown in Figure 8.1. Notice that only that only 3' extensions are practical. When using restriction enzymes to cut the vector, the only ones that will be compatable (besides blunt, but if I wanted blunt cloning I wouldn't be screwing around with this protocol in the first place) are those that leave a 3' overhang and have a T at the last cutting position (see Figure 8.1 for a list of potential enzymes). With USER cloning just replace this last T with a U and that provide the cut site for the USER enzyme mix. U in this case is uridine (the deoxyribose form of uracil). Most oligo synthesis companies will stick a U in for you for a small fee (e.g. IDT DNA).

An important point to consider with this approach is that most proof-reading Taqs are not compatable with primers that contain a U (it stalls the polymerase). The stratagene Pfu C_x is supposed to not have this problem. Otherwise, a non-proofreading taq will work, albeit with more errors.

8.2 Plan

1. try cloning a gene using USER and a digested pNEB193 vector (does adding extra base pairs to the primers help with the cutting? test...)
2. try cloning a gene using USER and a taq amplified pNEB193 vector with USER ends (i.e. completely restriction enzyme free)

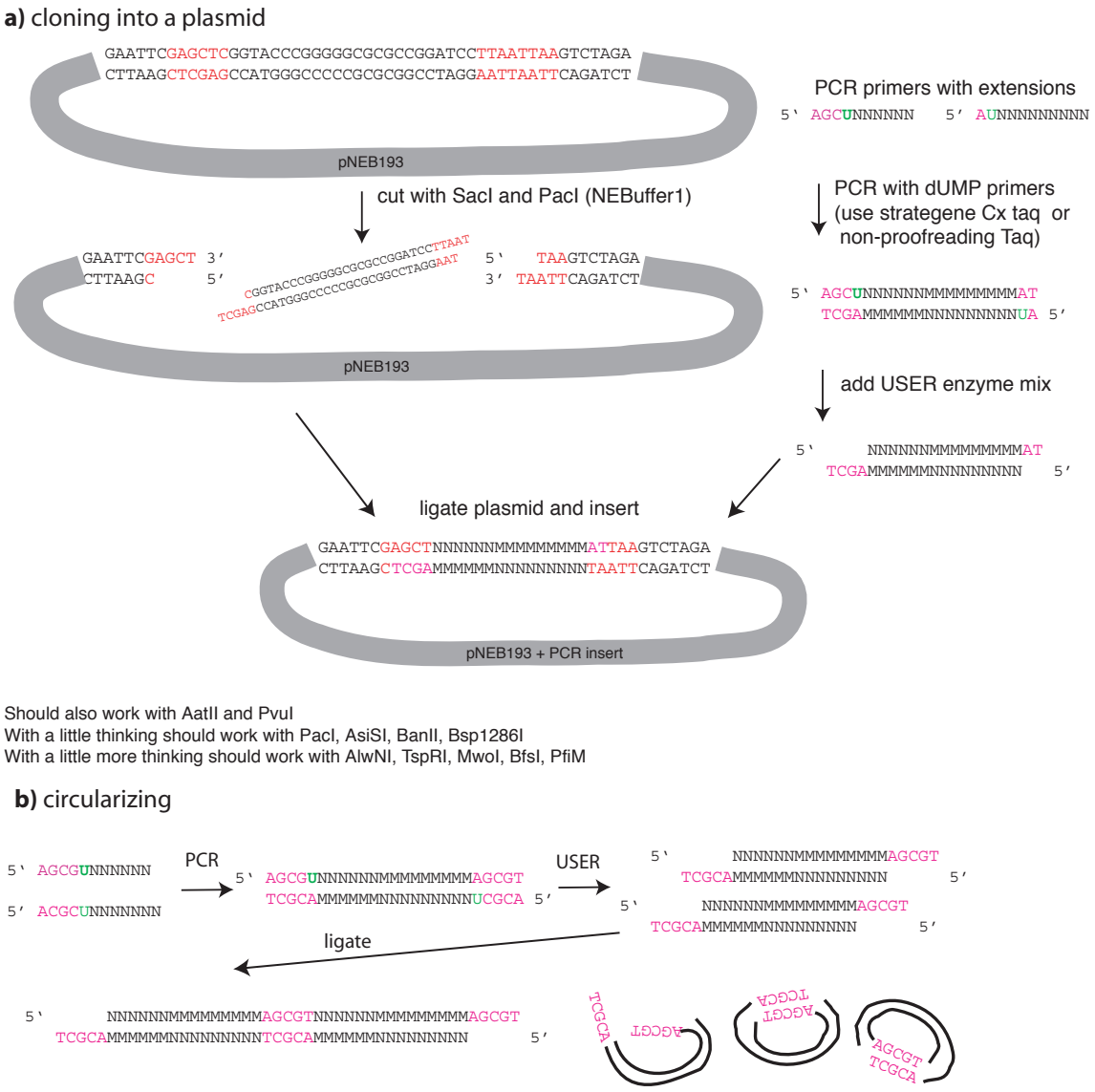


Figure 8.1: Example of USER cloning using SalI, PacI, and USER primers.

8.3 Cloning araB

As an initial test of the USER cloning strategy, I'm going to try and clone araB from the genome into pNEB193. I choose pNEB193, because it has a SalI and a PacI site, so it is possible to try directional cloning. I'd prefer to have a vector where I could clone using some combination of AatII, PvuI, and SacI as they all leave 4-bp overhangs, but NEB doesn't have such a vector and I didn't feel like digging around forever to find such a vector. As it stands, I think having the U only 2-bp from the end creates a bigger challenge anyways, so if it works I'll be pretty confident that it'll work with the other 4-bp overhang cutters too.

I choose araB, because it's pretty long (1701 bp), and it's also the first gene I could think of on the top of my head. The length make it suitable to try some circularization ideas later on and still be able to visualize the results easily on agarose gels. AraB has a AgeI site inside of it that I can use

to digest with after it is circularized.

```
> EG10053 araB (complement(70048..68348)) E. coli
atgGCGATTG CAATTGGCCT CGATTTTGGC AGTGATTCTG TGCGAGCTTT GGCGGTGGAC
TGCGCTACCG GTGAAGAGAT CGCCACCAGC GTAGAGTGGT ATCCCCGTTG GCAGAAAGGG
CAATTTTGTG ATGCCCCGAA TAACCAGTTC CGTCATCATC CGCGTGACTA CATTGAGTCA
ATGGAAGCGG CACTGAAAAC CGTGCTTGCA GAGCTTAGCG TCGAACAGCG CGCAGCTGTG
GTCGGGATTG GCGTTGACAG TACCGGCTCG ACGCCCGCAC CGATTGATGC CGACGGAAAC
GTGCTGGCGC TGCGCCCGGA GTTTGCCGAA AACCCGAACG CGATGTTCGT ATTGTGGAAA
GACCACACTG CGGTTGAAGA AGCGGAAGAG ATTACCCGTT TGTGCCACGC GCCGGGCAAC
GTTGACTACT CCCGCTACAT TGGTGGTATT TATTCCAGCG AATGGTTCTG GGCAAAAATC
CTGCATGTGA CTCGCCAGGA CAGCGCCGTG GCGCAATCTG CCGCATCGTG GATTGAGCTG
TGCGACTGGG TGCCAGCTCT GCTTTCCGGT ACCACCCGCC CGCAGGATAT TCGTCGCGGA
CGTTGCAGCG CCGGGCATAA ATCTCTGTGG CACGAAAGCT GGGGCGGCCT GCCGCCAGCC
AGTTTCTTTG ATGAGCTGGA CCCGATCCTC AATCGCCATT TGCCTTCCCC GCTGTTCACT
GACACTTGGA CTGCCGATAT TCCGGTGGGC ACCTTATGCC CGGAATGGGC GCAGCGTCTC
GGCCTGCCTG AAAGCGTGGT GATTTCCGGC GGCGCGTTTG ACTGCCATAT GGGCGCAGTT
GGCGCAGGCG CACAGCCTAA CGCACTGGTA AAAGTTATCG GTACTTCCAC CTGCGACATT
CTGATTGCCG ACAAACAGAG CGTTGGCGAG CGGGCAGTTA AAGGTATTTG CGGTCAGGTT
GATGGCAGCG TGGTGCCTGG ATTTATCGGT CTGGAAGCAG GCCAATCGGC GTTTGGTGAT
ATCTACGCCT GGTTTGGTCG CGTACTCGGC TGGCCGCTGG AACAGCTTGC CGCCCAGCAT
CCGGAACTGA AAACGCAAAT CAACGCCAGC CAGAAACAAC TGCTTCCGGC GCTGACCGAA
GCATGGGCCA AAAATCCGTC TCTGGATCAC CTGCCGGTGG TGCTCGACTG GTTTAACGGC
CGCCGCACAC CGAACGCTAA CCAACGCCTG AAAGGGGTGA TTACCGATCT TAACCTCGCT
ACCGACGCTC CGCTGCTGTT CGGCGGTTTG ATTGCTGCCA CCGCCTTTGG CGCACGCGCA
ATCATGGAGT GCTTTACCGA TCAGGGGATC GCCGTTAATA ACGTGATGGC ACTGGGCGGC
ATCGCGCGGA AAAACCAGGT CATTATGCAG GCCTGCTGCG ACGTGCTGAA TCGCCCGCTG
CAAATTGTTG CCTCTGACCA GTGCTGTGCG CTCGGTGCGG CGATTTTTGC TGCCGTCGCC
GCGAAAGTGC ACGCAGACAT CCCATCAGCT CAGCAAAAAA TGGCCAGTGC GGTAGAGAAA
ACCCTGCAAC CGTGCAGCGA GCAGGCACAA CGCTTTGAAC AGCTTTATCG CCGCTATCAG
CAATGGGCGA TGAGCGCCGA ACAACACTAT CTTCCAACTT CCGCCCCGGC ACAGGCTGCC
CAGGCCGTTG CGACTCTATA A

Forward
ATG GCG ATT GCA ATT GGC CTC G
MT: 61.1C

Forward + Adaptor
AGCU ATG GCG ATT GCA ATT GGC CTC G
MT: 63.6C

Reverse
TTA TAG AGT CGC AAC GGC CTG GG
MT: 60.9C

Reverse + Adaptor
AU TTA TAG AGT CGC AAC GGC CTG GG
MT: 60.9C
```

8.3.1 PCR with uradine primers, Linearization of pNEB193

Wed Nov 1 15:27:23 EST 2006

PCR with uradine primers

Wed Nov 1 15:27:27 EST 2006

The uridine araC primers were reconstituted to 100 mM. A primer mix at 10 mM was made with the forward and reverse primers (all primer stocks were made with TE).

The PCR was performed according to the Stratagene Cx manual which recommended an annealing temperature 5C below the lowest MT of the two primers (this ended up being 55C). The rxn was: 500 nM each primer, 1 μl dNTP (200 uM final), 1 μl PfuCx polymerase (2.5U), 40 μl H_2O , and appx 100 ng genomic DNA.

This is a true hot-start polymerase. Hot-start was 2 min at 95C. Extension was 1 minute. Polish was 10 min at 72C. I ran 30 cycles. 10 μl of the PCR product was run on an 1% μl agarose gel (see Figure 8.3).

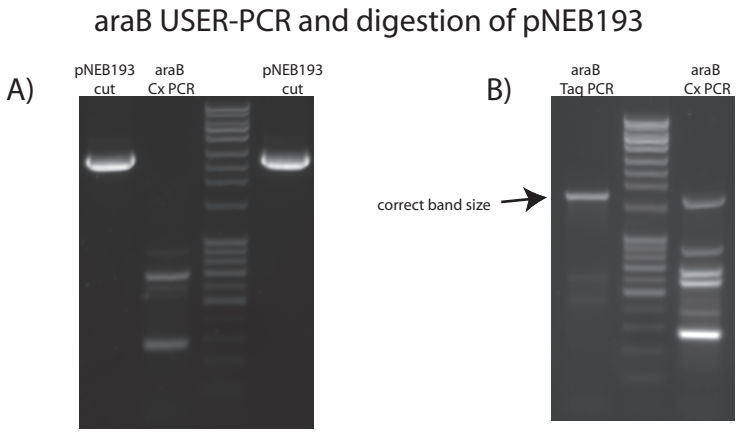


Figure 8.2: Digested pNEB193 and araB PCR with uridine

Brief Conclusions: The PCR definitely didn't yield the correct band of 1700 bp (Figure 8.3A). I'm not sure what those bands are? I'm going to try the PCR with a non-proofreading Taq.

Linearization of pNEB193

Nov 1, 2006

pNEB193 was ordered from NEB. I haven't made a freezer stock yet, so I'll just use their plasmid until I clone it and make my own.

I linearized the plasmid with blunt cutter SmaI. This will reduce any background transformations because the blunt ends shouldn't religate quickly. I also wanted to look at the plasmid one a gel one time before I used it.

The digestion was: 4 μl of pNEB193 (2 μg), 1 μl NEBuffer4, 4.5 μl H_2O , 0.5 μl SmaI. Digested for 10 minutes at RT. Heat deactivated 20 minutes at 65C. I ran 2.5 μl in each of the pNEB193 lanes in Figure 8.3A.

Brief Conclusions: Linearization worked well, I don't see any other bands besides the linearize plasmid in Figure 8.3.

PCR with uradine primers and non error-correcting Taq

Nov 2, 2006

I want to determine if it is the PfuCx taq that is messing things up with the Uradine containing araB primers. I ran the following reactions:

Taq rxn: 0.5 genomic (125 ng), 2.0 μl primer mix, 25 μl NEB Taq master mix, 22.5 μl H_2O

PfuCx rxn: 0.5 genomic (125 ng), 2.0 μl primer mix, 1 μl dNTP, 5 μl PfuCx buffer, 40.5 μl H_2O

I ran 10 μl of each on a 1% agarose gel (see Figure 8.3B).

Brief Conclusions: The Taq PCR definitely got the correct band of 1700 bp (Figure 8.3B), but it also got a lot of other junk. The PfuCx got closer to the right band size (though it might be 100 bp short?), but it also got even more junk than the normal Taq. I think my strategy for now is to order new primers for different genes (I'll return to araB later depending on how the new test works). I want to try lrp (from genomic) and mCherry protein (from plasmid). I'm going to order two sets of primers for each gene: 1 with U's and the other with T's. Hopefully this will clear up what's going on. I'm going to order these from invitrogen which allows a smaller yield and it quite a bit cheaper than IDT for these Uridine based oligos. I also want to get the primers with a higher melting temperature than last time (which was 60C, bump it up to 65C this time).

8.4 Trying with mCherry and lrp

Nov 7, 2006

Didn't have much luck USER cloning araB. I want to see if it a problem with IDT, araB, melting temperature, PfuCx, or if it just isn't going to work. I'm going to try one plasmid based gene (should be easier to amplify) and one genomic gene.

8.4.1 mCherry and lrp USER primer design

```
mCherry forward
ATG GTG AGC AAG GGC GAG GAG GAT AAC ATG GCC
33bp 68C

mCherry forward + adaptor
AGCU ATG GTG AGC AAG GGC GAG GAG GAT AAC ATG GCC
37bp 71C


mCherry reverse
TTA CTT GTA CAG CTC GTC CAT GCC GCC GGT GGA G
34bp 70C

mCherry reverse + adaptor
AU TTA CTT GTA CAG CTC GTC CAT GCC GCC GGT GGA G
36bp 70C
```

```
lrp forward
     ATG GTA GAT AGC AAG AAG CGC CCT GGC AAA GAT C
34bp 68C

lrp forward + adaptor
AGCU ATG GTA GAT AGC AAG AAG CGC CCT GGC AAA GAT C
38bp 70C

lrp reverse
TTA GCG CGT CTT AAT AAC CAG ACG ATT ACT CTG CTT GA
38bp 68C

lrp reverse + adaptor
AU TTA GCG CGT CTT AAT AAC CAG ACG ATT ACT CTG CTT GA
40bp 68C
```

This time I ordered the primers from invitrogen. They synthesize the lower amounts (25nM) for U containing primers, so it works out much cheaper ($12 a primer for these long ones with U, 5 dollars for the ones with a T). I've seen other people use the ones from Invitrogen, so hopefully these will work better than the ones form IDT.

8.4.2 USER PCR of mCherry and lrp

Resuspend primers at 100 μM . Then make a mix of 10 μM . Use 2 μl per 50 μl rxn for 400 nM final conc. I diluted the mCherry miniprep from section 10.1.3 page 390 to be 40 ng/μl .

Taq rxn: 0.5 genomic (125 ng), 2.0 μl primer mix, 25 μl NEB Taq master mix, 22.5 μl H_2O

PfuCx rxn: 0.5 genomic (125 ng), 2.0 μl primer mix, 1 μl dNTP, 5 μl PfuCx buffer, 40.5 μl H_2O

Taq cherry rxn: 0.5 plasmid (20 ng), 2.0 μl primer mix, 25 μl NEB Taq master mix, 22.5 μl H_2O

PfuCx cherry rxn: 0.5 plasmid (20 ng), 2.0 μl primer mix, 1 μl dNTP, 5 μl PfuCx buffer, 40.5 μl H_2O

Thu Nov 9 13:03:09 EST 2006

Actually, I just opened the envelope and noticed I only recieved the nonUSER primers. I guess it'll take them a little longer to make the ones with the U's. Save this exp for another day....

Mon Nov 13 19:06:52 EST 2006

Got the primers, running the PCR rxns overnight.

Wed Nov 15 14:40:23 EST 2006

I ran the PCR rxns on a 1.5% agarose gel.

I quantified the PCR yields from the gel using the using the Versadoc software:

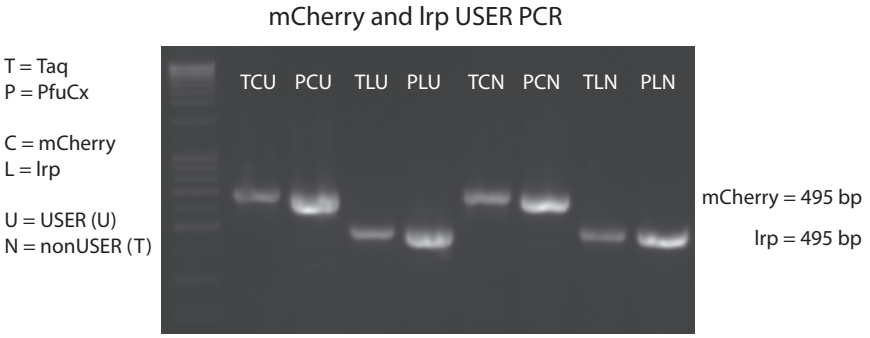


Figure 8.3: lrp and mCherry PCR with and without uridine using PfuCx and standard Taq. Gel is 1.5% run for 50 minutes at 100V.

	Linear extrapolation (per μl)	**Point-to-point extrapolation (per μl)**
TCU	223 ng (28 ng/μl)	93 (12 ng/μl)
PCU	382 ng (48 ng/μl)	143 (18 ng/μl)
TLU	171 ng (21 ng/μl)	77 (10 ng/μl)
PLU	351 ng (44 ng/μl)	133 (17 ng/μl)
TCN	218 ng (27 ng/μl)	91 (11 ng/μl)
PCN	386 ng (48 ng/μl)	144 (18 ng/μl)
TLN	163 ng (20 ng/μl)	74 (9 ng/μl)
PLN	269 ng (34 ng/μl)	107 (13 ng/μl)

My gut feeling is that these are underestimates. I'll use the linear extrapolations for the USER cloning since these are always higher.

Brief Conclusions: Wed Nov 15 15:51:16 EST 2006

It looks like everything is working just right with these Invitrogen primers. I don't know if it is the company, the increased length/mt, or the gene. I can now trying to clone these guys in using the USER enzyme. If the cloning step works, maybe I'll try to take one stab at figuring out why the araB didn't work (i.e. try invitrogen same primers and longer primers).

8.4.3 USER cloning of mCherry and lrp

Here we go...

preparing the pNEB193 vector

Thu Nov 16, 2006

I linearized the pNEB193 vector using SmaI in the following reaction: 4 μl pNEB193 (2 μg), 0.5 SmaI, 1 μl Buffer4, 4.5 μl H_2O . I incubated at RT for 30 minutes and deactivated at 65C for 20 minutes. This blunt-linearization was to prevent false positives in the ligation and to make the sticky ended inserts so short that they'd be easy to purify with a PCR purification (Qiagen).

After linearization, I added 1 μl Buffer1, 2 μl BSA, 6 μl H_2O , 0.5 μl PacI, and 0.5 μl SacI and incubated at 37C for 45 minutes followed by at 20 minute deactivation at 65C. The vector was cleaned up with a Qiagen PCR purification kit and eluted into 30 μl of EB buffer.

Sample	DNA (ng/ul)	260/280	260/230	total yield
cut pNEB193	46.0	1.78	1.62	1.26 μg

USER, ligation, transformation

Thu Nov 16, 2006

I pretty much just followed the instructions from the USER manual on the NEB website. I used their suggestion of adding ligase however, because my overhangs are much smaller than the ones used for the product that they sell.

20 ng vector, 10 μl PCR, 1 μl USER enzyme: incubate 15 minutes at 37 C.

add 1 μl T4 DNA ligase buffer, add 1 μl T4 DNA ligase: incubate at RT for 15 minutes.

I transformed 2 μl of each ligation/USER reaction into DH5α. I plated 50 μl . I ran 5 total reactions: (using the nomenclature from the previous section) PUC, PUL, PNC, PNL, - control (plasmid only to make sure the blunt digestion and Qiagen cleanup efficiently removed the insert). The plates contained x-gal and IPTG for blue/white screening.

Brief Conclusions: how many colonies, did it work? are they white? if so pick for miniprep, pcr check and sequencing

Insert check of USER clones

I picked 8 clones, grew them overnight, miniprepped them, and checked the inserts by PCR using the M13 primers. The mCherry clone overnights were bright red, indicating that the correct gene was there.

Nov 29, 2006 I ran a PCR on the minipreps and ran 10 μl of the PCR rxn onto an agarose gel (Figure 8.4).

Figure 8.4: Insert check for lrp and mCherry clones with and without uridine using PfuCx and standard Taq. Gel is 1.0% run for 50 minutes at 100V.

Brief Conclusions: The colored proteins show that the right thing is in there, even for the genes were I didn't have sticky ends. None of the inserts really amplified (Figure 8.4). All you can really see are the plasmids and the primers. I'm not sure what's going on.

Chapter 9

Dabbling in synthetic biology

THIS CHAPTER/PROJECT IS NOT ACTIVE

See the first figure for a pictural overview. See the last page for a summary of the results.

Make toggle with components from systems biology predictions.

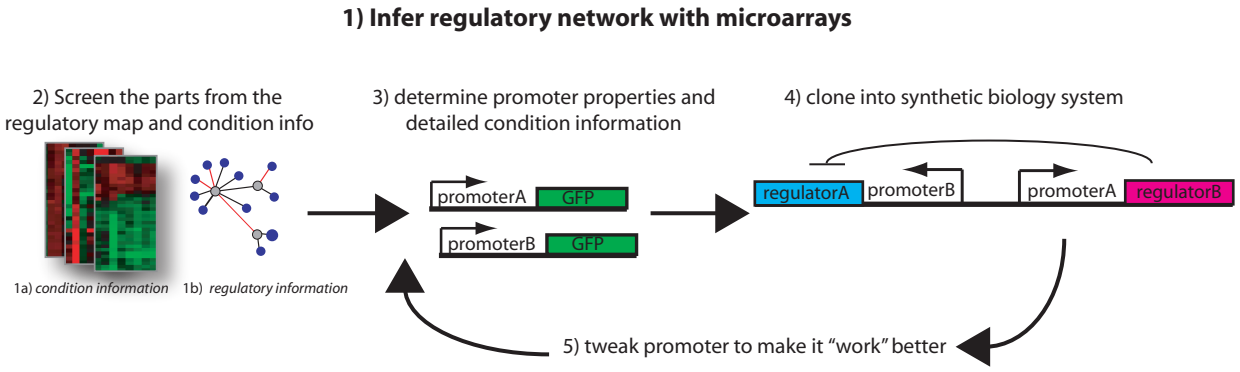


Figure 9.1: The basic schema for applying network inference algorithms to synthetic biology problems

9.1 testing pIKE 107

can I switch it? need to order tetracycline

9.2 making a pyruvate / lactose regulated switch

Instead of tetracycline the new switch (if it works) will use pyruvate which derepresses PdhR protein (see Figure 9.2).

9.2.1 Cloning in ndh and pdhR promoters

Thu Sep 13, 2006

Ilaria already has working protocols for sticking these promoters into vectors, so for the most part I'm just copying what she's done already but with new genes/promoters. On the pIKE vector, I'm switching out the pLtet promoter and the tetR genes with pdhR/ndh promoters and pdhR gene respectively. The pdhR promoter is a known target of the pdhR repressor; ndh is a new one that I verified in the chip studies. In this section I'm trying to swap the pLtet promoter with a pdhR one and a ndh one (i.e. to make 2 separate vectors). Then I'll add pdhR to each and see if I can make a pyruvate/IPTG switch [1].

PCR the gene from the genome

I amplified each promoter from genomic DNA[2]. The primer pairs each added a site for SphI at the 5' end and AatII at the 3' end.

[1] I realize it is more parsimonious to add pdhR gene and then the two promoters, but the pdhR gene has a cutter in the middle of if that is needed to clone the two promoters.

[2] I considered the *promoter* to be the first 200 bp upstream of the ATG start codon

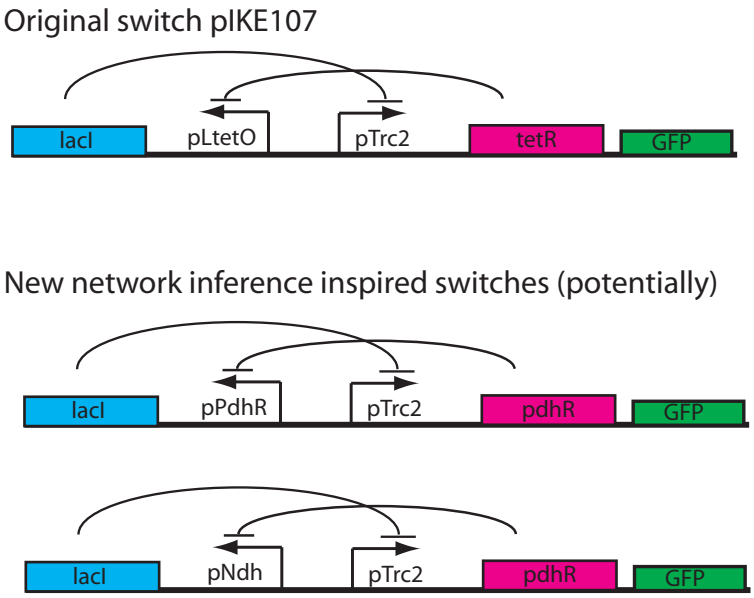


Figure 9.2: The potential toggle built using network inference information will replace the tetR gene from Tim Gardner's pIKE107 toggle with pdhR protein. One of the new toggles will use the known repressor site of the PdhR protein that resides prior to the pdhR-aceE operon. The other new toggle will use the new predicted target of pdhR, verified by ChIP: ndh.

I ran 30 μl PCR reactions using the Easy-A master mix [Stratagene]. The MT for the first 5 cycles was 56 and was then ramped up to 62 for the last 25 cycles (once there were enough sequences with the extra 10 bp or so added by my primers). Yields were unimpressive and post PCR cleanup DNA was pretty dirty. But Ilaria's done this 100x and said all of these things were normal:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**total yield**
ndh promoter	45.1	1.74	1.23	1.4 μg
pdhR promoter	43.4	1.86	1.12	1.3 μg
pdhR gene	62.1	1.88	2.47	1.9 μg

Digesting the vector and the inserts

Sep 13, 2006

I minipreped pIKE107 from an overnite culture. Yield:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**total yield**
pIKE107	139.9	2.02	2.26	7.0 μg

I digested 15 μl of vector for 1 hour in a 20 μl double digest with 0.5 μl of each cutter (AatII SphI). N μl of the PCR products were also digested in a similar manner, but for 45 minutes.

The PCR'd insert digestions were purified with a Qiagen PCR purification kit. Yields were crappy as was the nanodrop spec reading. Ilaria said this was all normal, not to worry.

Sample	**DNA (ng/ul)**	**260/280**	**260/230**	**total yield**
digested, PCR purified ndh promoter	28.9	1.71	0.75	867 ng
digested, PCR purified pdhR promoter	7.7	1.53	41.78	231 ng

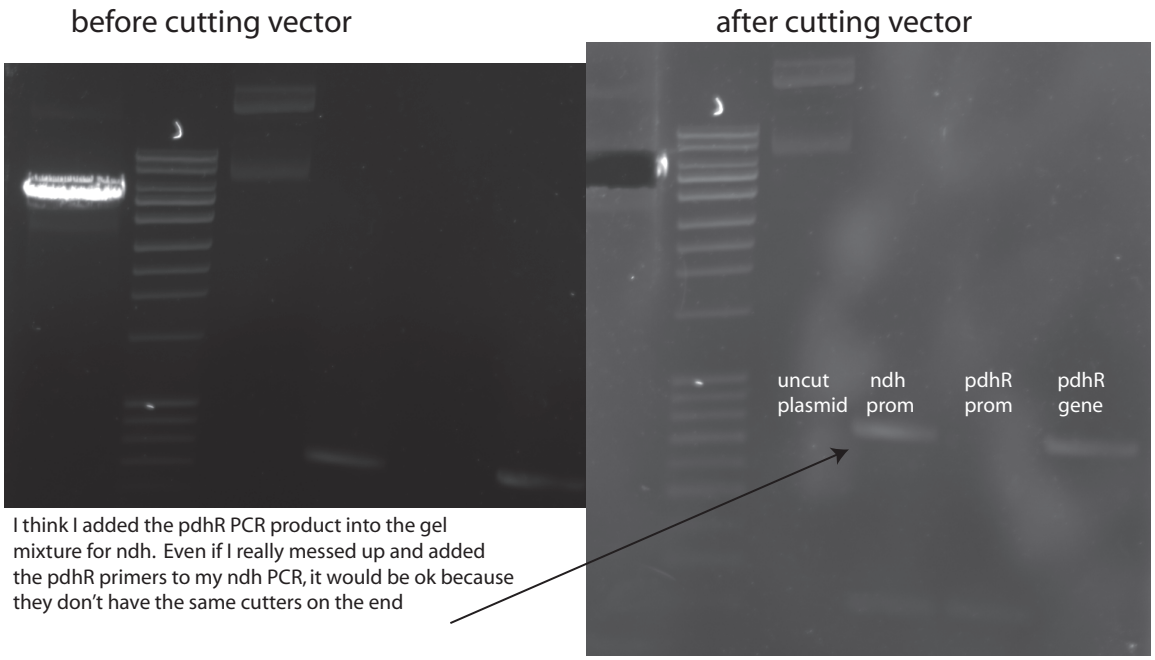


Figure 9.3: 1% agarose run for 75 minutes at 120 V. I show the gel after the chunk was remove for purification by razor blade because it was hard to see the PCR fragments while it was there.

The vector was gel purified (to get rid of the insert). Yield and spec readings for this also sucked. Again Ilaria said it was normal.

See Figure 9.3 for the inserts (uncut) and cut vector on a gel. I ran the pdhR gene on the gel too just to test the primers.

Sample	DNA (ng/ul)	260/280	260/230	total yield
digested, gel purified pIKE107	16.2	1.90	0.04	366 ng

Ligation and transformation of ndh and pdhR promoters

I made 3 ligations. ndh prom + vector, pdhR prom + vector, vector alone (- control). For each I used 10 μl of vector. For ndh prom and pdhR prom, I used 2 μl of the cleaned PCR product. The ligation was run at 16 C for 30 minutes. Followed by the standard ligation proceedure. I plated 75 μl of each transformation (too much).

There were MANY colonies for the two samples and very few on the negative control. I picked 3 colonies for each of the two promoters.

Checking for the proper insert size

I ran PCR reactions with a melting temperature of 57 C for each of the 6 miniprepped samples using 0.5 μl plasmid and 1 μl of 5 uM primer per rxn.

Brief Conclusions: Each plasmid type had two out of three colonies with the correct insert length (see Figure 9.4.

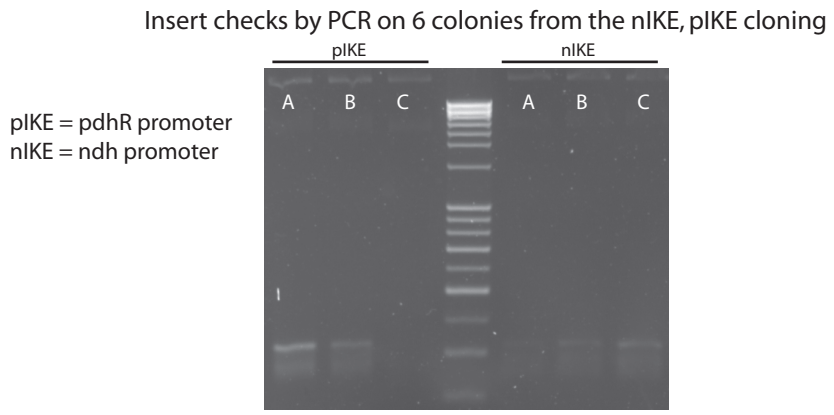


Figure 9.4: 1.5% agarose. I'm not sure why there is a smear below the correct sized band.

Sequencing the inserts

The four plasmids with the insert lengths of the correct size were spec'd and sent out for sequencing.

Sample	DNA (ng/ul)	260/280	260/230	total yield	Primer Orientation for sequencing	Sequence Correct
pIKE A	109.1	1.96	2.19	5.5 μg	F	Y
pIKE B	135.4	1.95	2.19	6.8 μg	R	
nIKE B	140.5	1.99	2.21	7.0 μg	F	Y
nIKE C	167.0	2.02	2.20	8.4 μg	R	

All of the sequence data and chromatagraphs from Agencourt can be found here.

9.2.2 Cloning the pdhR gene into the pdhR-prom and ndh-prom vectors

Thur Sep 21, 2006

I digested 15 μl (appx 2.3 μg) of pIKE-a and nIKE-b with PstI, AgeI, buffer I, and BSA for 1 hr at 37 C followed by 80 C heat inactivation for 20 minutes. In the same manner, I digested 15 μl of pdhR gene PCR product to which I had added the appropriate restriction sites. The digested pdhR gene was cleaned with a Qiagen PCR purification kit and spec'd:

Sample	DNA (ng/ul)	260/280	260/230	total yield
pdhR gene	18.6	2.49	5.14	558 ng

All 20 μl of the two plasmid digestions were run on a prestained Sybr Safe gel, imaged with the versadoc (using UV) for 1.4 seconds. And cut with a razor blade while viewing on the transilluminator (i.e. no UV).[3] The gel slices were cleaned using a Qiagen gel cleanup kit (see Figure 9.5).

10 μl [4] of each cleaned gel slice was used in the ligation reaction. For each ligation, 2 μl of digested, purified pdhR gene was used. A negative control was used for pIKE-a and nIKE-b, consisting of the cut plasmid without insert. Ligations were done for 30 minutes at 16 C with no ligase heat deactivation.

The transformation was done as normal except I used 350 μl of SOC (normally I use 250 μl) and I only plated 50 μl .

[3] This is the first time I've used Sybr Safe. It is very clear on the transilluminator, much better than EtBr.
[4] from the total 30 μl elution volume of the Qiagen kit

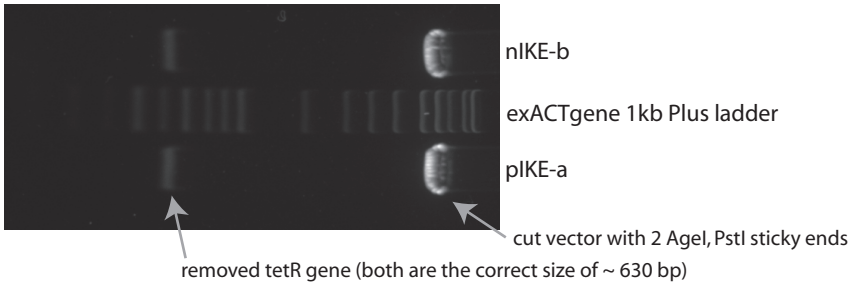


Figure 9.5: 1% agarose run for 85 minutes at 90 V.

Brief Conclusions: There weren't very many colonies. Next time, I think I'll continue adding 350 μl but I'll plate 75 μl instead of 50 μl . There were enough colonies on each plate (5-20 or so) that I'm pretty sure they'll be some good ones. One negative control plate had one colony. The other had zero.

Checking the pdhR inserts in nIKE and pIKE

Sat Sep 23, 2006

Miniprepped 6 colonies (3 nIKE and 3 pIKE).

Sun Sep 24, 2006

I spec'd the 6 minipreps:

Sample	DNA (ng/ul)	260/280	260/230	total yield
nIKE-pdhR A	167.9	2.02	2.16	8.4 μg
nIKE-pdhR B	160.9	2.00	2.18	8.0 μg
nIKE-pdhR C	182.7	2.02	2.18	9.1 μg
pIKE-pdhR A	175.9	2.00	2.20	8.8 μg
pIKE-pdhR B	117.7	2.01	2.21	5.9 μg
pIKE-pdhR C	103.0	1.97	2.21	5.2 μg

and digested 8 μl the colonies with AgeI and PstI [5].

Brief Conclusions: All six of the checked inserts were of the correct size (see Figure 9.6). I'm going to send 2 of each for sequencing as soon as the primers arrive. For now I think I'll use use them assuming they're correct to save time (ok, I know it will be a waste of time if I discover the inserts are actually wrong.)

Brief Update *Sun Oct 29 17:32:10 EST 2006*: ndh-pdhR-a and pdh-pdhR-b were both fine (sequenced onthe 28th of September, 2006. pdh-pdhR-a was a little questionable raw sequencing data.

[5] I would have preferred to check them by PCR, but I worgot to buy the primers to check these things with

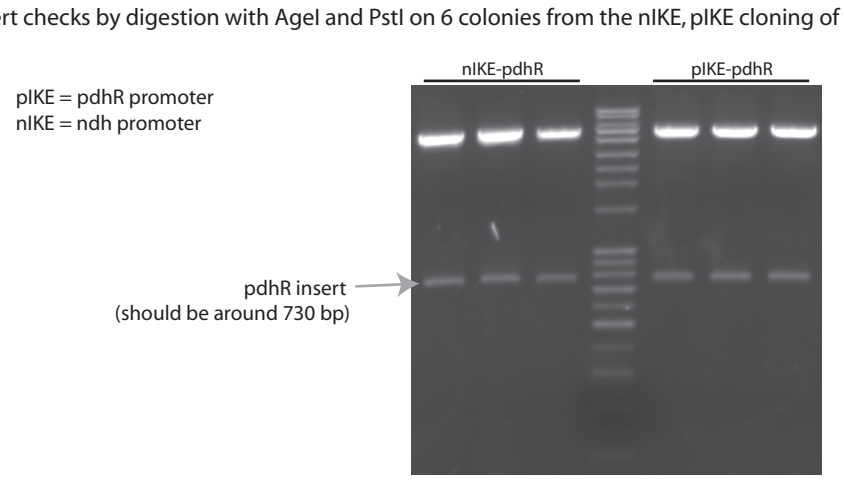


Figure 9.6: 1% agarose run for 55 minutes at 110 V.

9.2.3 Ilaria's testing of the individual promoters in the U. Alon plasmid

Ilaria screened a lot of different conditions. All of this stuff can be found in her notebook. I want only to show the promoter responses for the final conditions used in the working attempt at switching the toggle (shown below).

Ilaria uses the U. Alon plasmid (Figure 9.7) and a simple model to estimate gene expression from GFP time-series GFP measurements.

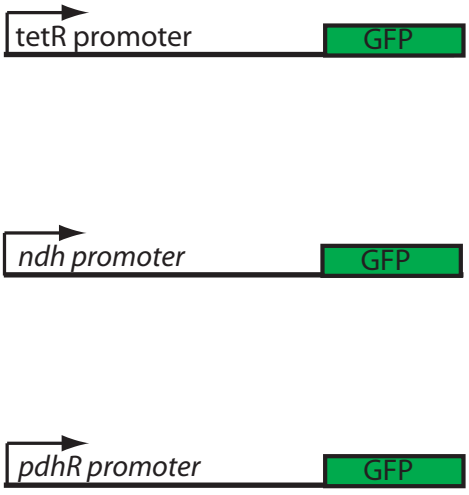


Figure 9.7: Promoters to be used for the toggle switch were tested individually across a variety of conditions using the Alon plasmid.

The estimates correspond pretty well to actually expression measurements obtained with qPCR (see Fig 9.8).

Ilaria tested the promoters across many conditions; the ones relevant to the pdhR toggle are shown in Figure 9.9. She has many others tested conditions. All of the conditions without saturating glucose have a double hump where the cells first eat all of the available nutrients (first hump) and

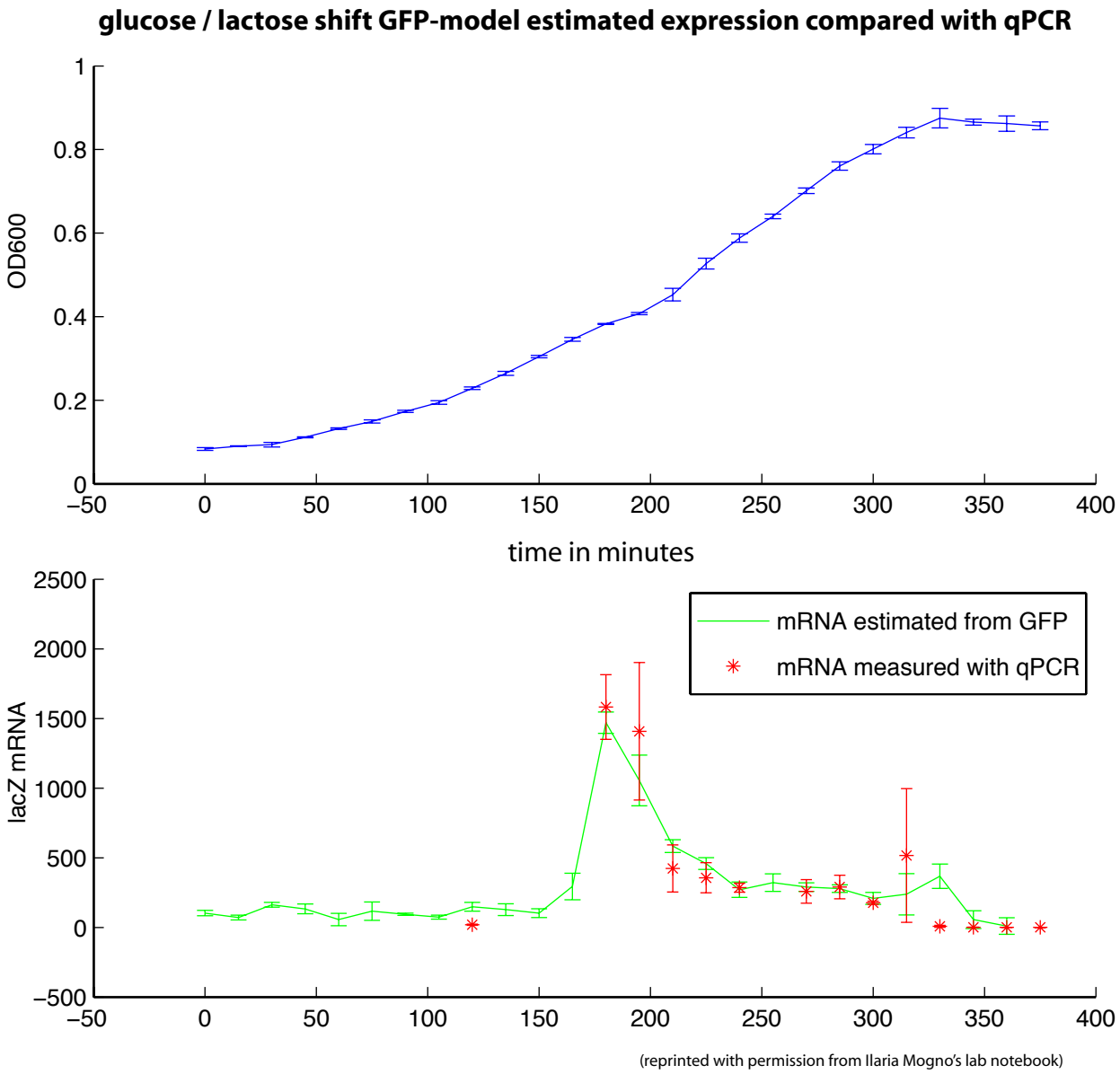


Figure 9.8: Gene expression of lacZ estimated from GFP measurements corresponds well with directly measured expression using RT-qPCR.

then finally eat the leftover pyruvate (second hump) (this is just the current guesstimate to why this happens. I felt the hump was causing the switch to continually go off with time so this high glucose way keeps a constant signal on the promoter (as long as you don't allow the cells to reach late stationary phase. One interesting result (not shown but in Ilaria's notebook). When I designed the pdhR promoter, I used the 200 bp upstream of the operon. When Ilaria made her initial pdhR promoter strain, she used 500 bp upstream. When she compared the 200bp to the 500bp, the 200bp promoter is around 2x stronger (image not shown see Ilaria's notebook or here Oct 2006 lab meeting slides).

Last, notice that the ndh promoter, although it was strongly enriched in the ChIP study only shows

a weak increase in expression when pdhR promoter is derepressed. This might have more to do with the strength of the -10 and -35 than with the binding efficiency of pdhR protein to the ndh promoter. It could also be that ndh needs an activator to express fully.

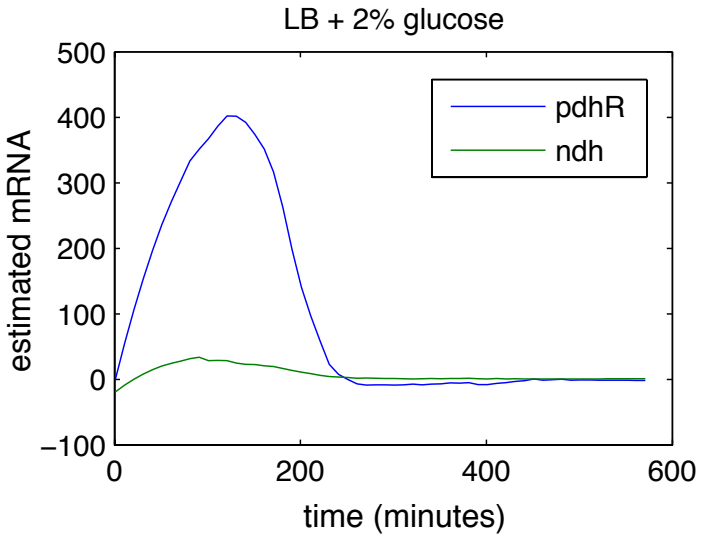


Figure 9.9: Gene expression estimates for ndh and pdhR promoters.

9.2.4 Testing the pdhR version of the toggle

Mon Sep 25 12:43:38 EDT 2006

I now have two new versions of the toggle, one with a ndh promoter and one with a pdhR promoter. Ilaria has already tested that the 200 bp pdhR operon responds in to pyruvate in her gfp system. I'm a little worried because the lac repressor is so strong. Perhaps I should've used tetR instead of lac? We'll see how this goes. I'm pretty sure it's not going to be a problem to get it off. It's keeping it on at a decent amount that I'm unsure of. Ilaria is also building a pLpdhR promoter that should have more gain to it. I really would like a pLndh.

the plan of action for testing the new potential switch

I'm going to toggle the two new pdhR based switches and the old pIKE107 that I've already tested in the plate reader and I know works (see section 9.1). Based off what I learned in the previous test of the toggle here's what I'm going to do to test these guys:

1. test nIKE-pdhR, pIKE-pdhR, nIKE, pIKE, and pIKE107
2. toggle all strains *off* overnite using appropriate chemical (1% pyruvate for the nIKE and pIKE based plasmids and 500 ng/ml aTc for pIKE107; use LB media
3. measure OD and fluorescence of the stationary phase culture
4. wash 2x in LB and dilute 1/200 in LB

5. measure OD and fluorescence until cells approach early stationary phase
6. wash 2x in LB and dilute 1/1000 in LB [6] with appropriate chemical to turn GFP *on* (IPTG or lactose, I'd like to try both)
7. measure OD and fluorescence of the stationary phase culture
8. wash 2x in LB and dilute 1/200 in LB
9. measure OD and fluorescence until cells approach early stationary phase

the plan I tried to take

The background fluorescence from LB was really messing things up. I tried to switch to a Davis minimal + casamino acids format. However, they just grew too slow.

the plan I actually took

I grew the cells in LB media with ampicillin. For the pdhR based plasmids, the LB also always had 2% glucose. Samples were taken every 45-120 minutes. For each sample the cells were washed 2x in Davis media with glucose. This removed the LB media and allowed small quantities of GFP to be measured free from the background fluorescent problems associated with LB. 200 μM IPTG was used to turn both switches on. The toggle on and toggle off states were both done overnite using a 1:1000 dilution. Prior to adding the appropriate switching chemical, the cells were washed 2x in LB. Only one sample was taken for these samples on the morning following the switching (i.e. I only get one datapoint and that point is taken at stationary phase).

For the stable on and stable off states, I sampled every 45-120 minutes until the cells reached early stationary phase. Low OD measurements, where GFP/OD was not accurately estimatable, were removed from the analysis. I turned Tim's pIKE107 as described in Tim's toggle paper in *Science*.

I expected that I would toggle the new toggle off with pyruvate and on with IPTG. However, this did not work. Using this method, the switch would not stably stay on. My guess for an explanation is that pretty much everything you can feed the bacteria (especially in rich media like LB) gets turned into pyruvate eventually by central metabolism. The pyruvate then turns off the switch. I noticed that this turning off appeared to roughly correspond to the peaks Ilaria saw witht the Uri Alon plasmid that yielded two humps in the expression level of ndh and pdhR promoters when grown on minimal media + pyruvate and casamino acids. Also like Tim's plasmid there was a basal level of expression even in the off state.

The method that actually worked was to allow the cells to reach late stationary phase when putting them into the off state with pyruvate + 2% glucose (by growing overnite and being lazy the next morning about getting started). Wash the cells 2x with Davis minamal and take your sweet time about it. The late stationary combined with washing with Davis ensures that gene expression is essentially halted and the *memory* manifest in the mRNA population degrades and is forgotten. If I then started the cells (diluting 1:500) in LB + glucose, the GFP was essentially off - hardly more than in a pUC19 strain that contained no GFP. To turn them on just wash 2x in LB dilute 1:1000 in LB + glucose and add 200 μM IPTG. To keep them stably on, don't allow the cells to reach late stationary phase, instead at late log phase (OD 0.8 or so) wash 2x in LB + glucose and dilute

[6] I use 1000 here because the cells have plenty of time to grow before the next morning

1:500 into LB + glucose, the switch should stay on. Note the ON state and the stable ON states are much lower with this LB + glucose that I could obtain with out having the glucose around. However, the off state is much lower with the LB + glucose too. And when using the LB + glucose, the toggle will stay in the two states stably.

These results can be seen in Figure 9.10. Not the y-axis is log scaled. Tim's pIKE1107 toggle has muc expression than the new toggles. However both types of toggle have about the same order of magnitude difference between their off and on states.

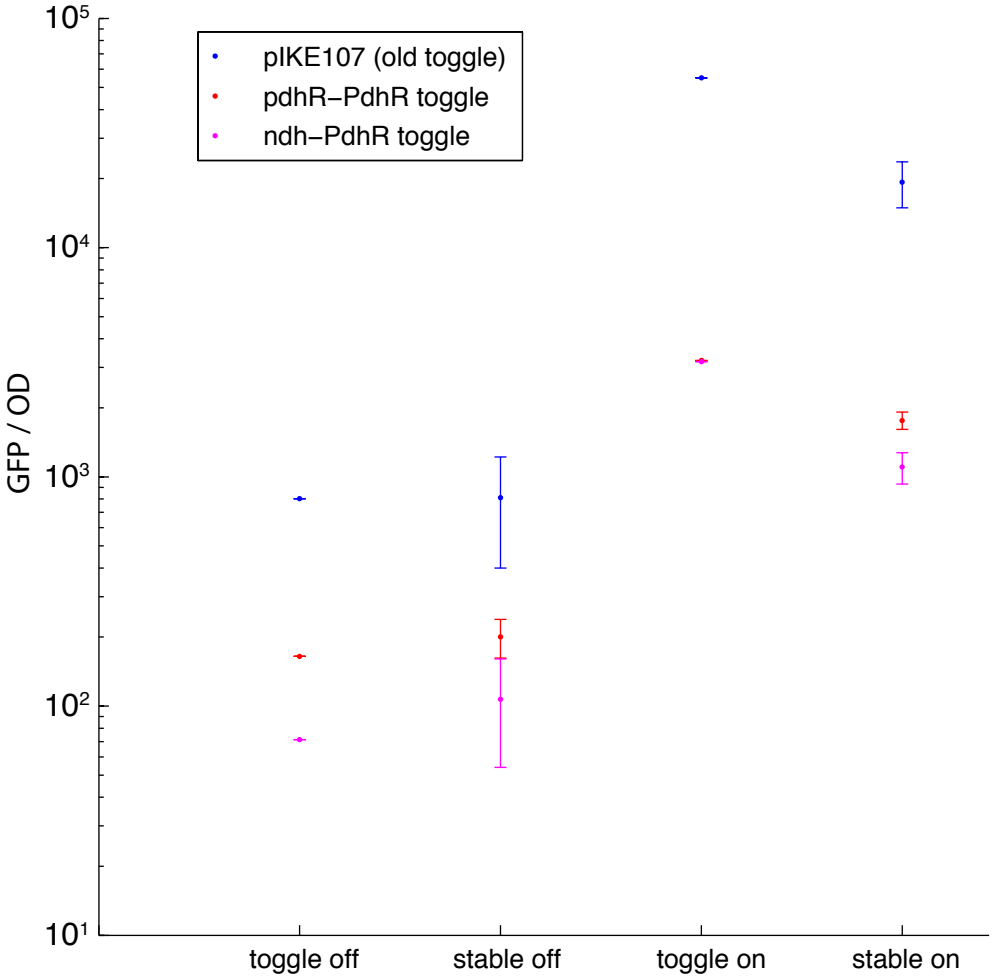


Figure 9.10: toggling the toggle notice the y-axis are on a log scale

Brief Conclusions: It seems like the toggle built using network inference results does stably toggle between two states. I presented these results at BMES in Chicago, Oct 2006. As a little part of a talk at the synthetic biology session. I used this to show how systems biology (and network inference in particular) can be used to do some synthetic biology (which currently reuses

the same 3-4 promoters over and over and over and over again [cI, lac, tet, ara]. Unfortunately, I only ran this test one time, I'm not completely confident in these results and would be much confident if I could repeat this a time or two. However, I don't really have plans to do so. Toggling this thing is very time-consuming and I'm moving on to a new side-project with Ilaria working towards measuring GFP and gene expression in single-cells combined with sequencing to look at how changes in promoters cause changes in gene expression and protein production.

Chapter 10

Miscellaneous Experiments

This chapter is for things that are too big to include in other chapters, but too little for their own chapters. This includes optimization of protocols in the appendix (where it isn't desirable to have all of the optimization info, I just want the best protocol there).

10.1 New color proteins from the Tsien lab

I requested 3 fluorscent proteins from the Tsien lab after reading their Nature Methods paper: *A guide to choosing fluorescent proteins*. Two of the colors, mCherry and mOrange, come from the paper: *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein*. The other protein mCitrine was published in *Reducing the Environmental Sensitivity of Yellow Fluorescent Protein*. This set really fills out our possibilities (before we only had GFP in the lab, then recently I got DsRed). I didn't request the proteins for an particular project. I thought I might attach the Cherry to the other side of the toggle[1]. But I have lots of other projects where multiple colors will be very useful. For the moment we have Red, Orange, Yellow, and Green. The only colors we're missing are a Cyan (Tsien lab doesn't have a good one) and a dark red / plum (Tsien lab has a decent, but dim one however I already requested my limit of 3 proteins; mCherry is pretty close to plum in excitation anyways).

I choose these three because they were the brightest and most separated, based on the info in *A guide to choosing fluorescent proteins*. GFP the other color we have in the lab is too close to Citrine to really add it as a fourth color. However, CyPet and Cerulean are cyan proteins that are still pretty bright (all of the blues are currently way too dim, not photostabile, and require UV excitation which damages DNA) and are pretty distant from mCitrine. Getting either of those two would allow 4 color imaging. If you are interested in three color, that's possible now using GFP, mOrange, and mCherry. However, you shouldn't use the wild-type GFP because it has a bimodal absorption pattern (395 and 475). EGFP with a single peak at 484 is a better option. A better three protein option is Cerulean (433), mCitrine (516), mCherry (587).

Figure 10.2 shows a couple images I took on the blue light transilluminator using the cheap lab digital camera. The cells are in Davis minimal media[2]. The second photo is the same as the first except I pelleted the cells to concentrate the proteins to one location and make them easier to see. The order of the second photo is Citrine, mCherry, mOrange.

[1] Tim's toggle is currently (Green / No Green) but there's no reason it couldn't be (Green / Red)

[2] LB fluoresces way to much in the yellow and orange area, making it hard to see the proteins in LB

10.1.1 Filter Selection

Below I list the filters recommended in the Tsien lab paper *A guide to choosing fluorescent proteins*. I also list the closest filters to the recommended ones that we have in our plate reader (which are the ones I actually use for the moment). The Single column refers to a sample with a single-protein. The multiple column refers to samples with multiple fluorescent proteins mixed together.

Protein	Best available in our plate reader	Tsien Single	Tsien Multiple
mCitrine	485/20, 528/20	490/30, 550/50	495/10, 525/20
mOrange	530/25, 590/35	525/20, 595/80	545/10, 575/25
mCherry	590/20, 645/40	560/20,640/100	585/20, 675/130

10.1.2 Archiving the colors

Oct 26, 2006

The samples came via standard US post on filter paper. I added 50-100 μl of TE (enough so that I could see some liquid at the bottom of the tube and let the filter paper sit in the TE for 30-60 minutes. I transformed 1 μl of the TE into DH5α cells.

10.1.3 Putting the colors in a strain where they'll express

Tue Oct 31 18:07:52 EST 2006

The proteins all have T7 promoters, so you get very little expression in DH5α cells (though I could still clearly see the cherry cells by eye). I'm putting the plasmids containing the proteins into BL21(DE3)pLyseE cells [Invitrogen] which have the T7 polymerase, so that I can mess around with the cells a little with the light microscope and plate reader to get a feel for them. The BL21(DE3)pLyseE has T7 lysozyme on a chloramphenicol maintained plasmid. The lysozyme serves to lower the basal expression from the T7 promoter (which is supposed to be extremely high). The cells also have the nice property that if you freeze thaw them a few times on ice, they lysozyme causes them to lyse themselves. This might have some biotechnology applications.

miniprepping the plasmids

Tue Oct 31 18:35:00 EST 2006 I grew the cells up last night and miniprepped them this morning:

Sample	DNA (ng/ul)	260/280	260/230	total yield
mCherry	352.4	1.92	2.23	17.6 μg
Citrine	302.4	1.94	2.16	15.1 μg
mOrange	49.6	1.89	2.07	2.5 μg

The mOrange is low because I messed up the elution of the miniprep. Instead of doing the final spin-down to remove excess ethanol, I added the EB elution buffer. I didn't want all of the ethanol in my sample. I added 750 μl ethanol in a hope that it would change the pH or at least precipitate the DNA a bit so that it wouldn't flow through the column. I then ran the EtOH through, and did a second spin to dry the residual etOH and finally added 50 μl EB buffer to elute. This killed my yield, but it's still enough DNA to do what I want. After I was finished, I realized it would've been much easier to elute the DNA + EB + ethanol and then EtOH precipitate the sample. Next time....

Cloning into BL21(DE3)pLyseE

Tue Oct 31 18:35:00 EST 2006

I cloned the cells according the Invitrogen protocol. I added chloramphenicol to some amp plates I already had and I plated 70 μl of the transformation.

Inducing expression

Wed Nov 1, 2006

IPTG induces the expression of T7 in the BL21(DE3)pLyseE cells. They say to use 0.5 mM in the Invitrogen manual. I picked one colony from each BL21(DE3)pLyseE transformation. I picked the same colony 2x, the first went into a falcon tube with LB and the appropriate antibiotics, the second went into a falcon tube with LB, appriopriate antibiotics, and 500 uM IPTG.

I was surprised to find that the samples with IPTG didn't really grow any noticable amount by the time the non-IPTG samples had reached stationary phase. I made a freezer stock of the stationary phase cultures (strains are in row 2 of the *main* box). I then reinoculated the IPTG samples using the left over stationary phase culture (1/7 dilution).

I saw a fairly immediate jump in the concentration of mCherry and especially Citrine. mOrange, which is supposed to be the most highly expressed protein, was not really detectable (Figure 10.1). However, the next morning, after overnite growth in IPTG, it was clear that the orange worked because the sample was very orange. The plate reader agreed with this observation 10.1. I read that the mOrange is very pH sensitive. It could be that the acidic conditions of LB was lowering the amount of fluorescent mOrange protein?

The time samples and the final overnite reading are available here in excel format. Notice also how the readings improve when using the autosensitivity adjustment (see Figure 10.1).

Brief Conclusions: Thu Nov 2 17:45:57 EST 2006

The plasmids certainly produce the correct colors (see Figure 10.2). The only thing that still worries me a little is why didn't the mOrange show up initial post-induction with IPTG? I learned that the autoscale feature can be very useful to make the detection readings more comparable with each other (Figure 10.1). Choosing the same sensitivity for every protein yielded much lower results for the high wavelength (red) readings - even for the LB blank. By autoscaling each well so that the LB blank is at value 100 the colors become much more comparable. For davis media, scaling the blank to 40 would probably be more appropriate.

10.2 Sephacryl column separation of DNA

I found this protocol: http://www.genome.ou.edu/protocol_book/protocol_partI.html *Fragment purification on Sephacryl S-500 spin columns* that described using a spin column stuff with S-500 Sephacryl to size-fractionate a sheared sample. Column size-fractionation is the method preferred for generating cDNA libraries. I guess it produces more ligatable DNA than gel-purification. I decided to give this protocol a whirl.

Here is the original protocol (just cut-and-paste from their website): DNA fragments larger than a few hundred base pairs can be separated from smaller fragments by chromatography on a size

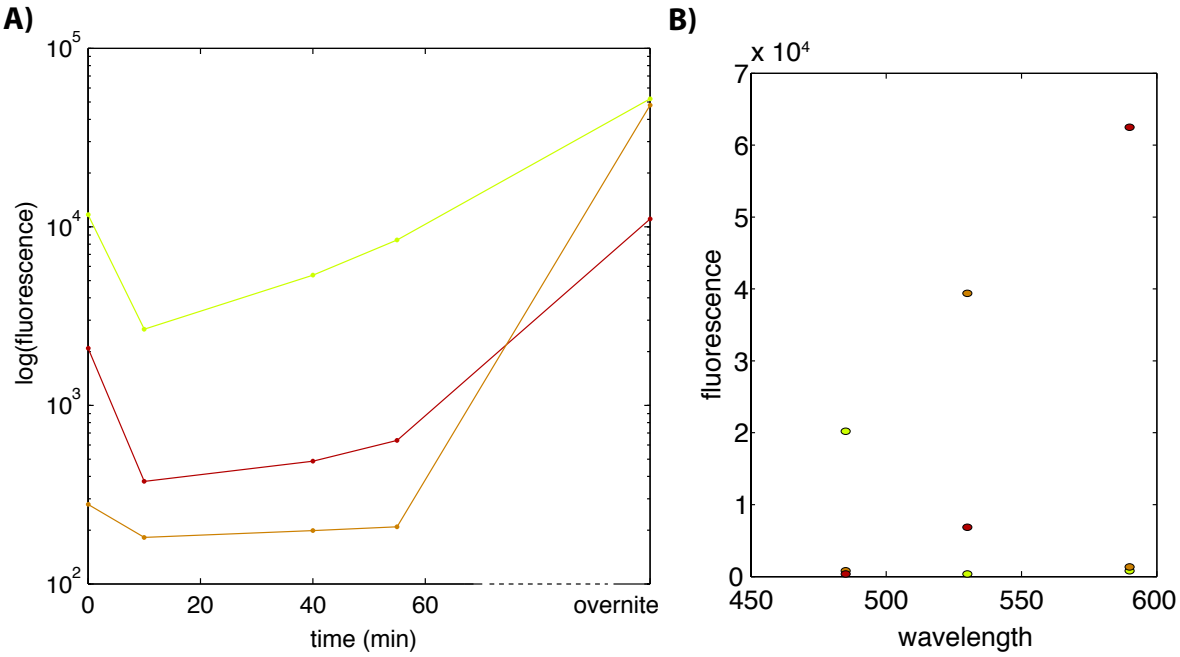


Figure 10.1: The fluorescence of the colors proteins from the Tsien lab was measured using the plate reader. The measurements after 1/7 incubation into 500 uM IPTG were done using a sensitivity of 50. The next day, I realized that using autoadjust was a way to get a nicer scaling. Initially (A), the more towards red you got, the lower the numbers were, even for the - control. The color separation is pretty good for the Citrine and mOrange. The color separation is *very* good for the mCherry.

exclusion column such as Sephacryl S-500. To simplify this procedure, the following mini-spin column method has been developed.

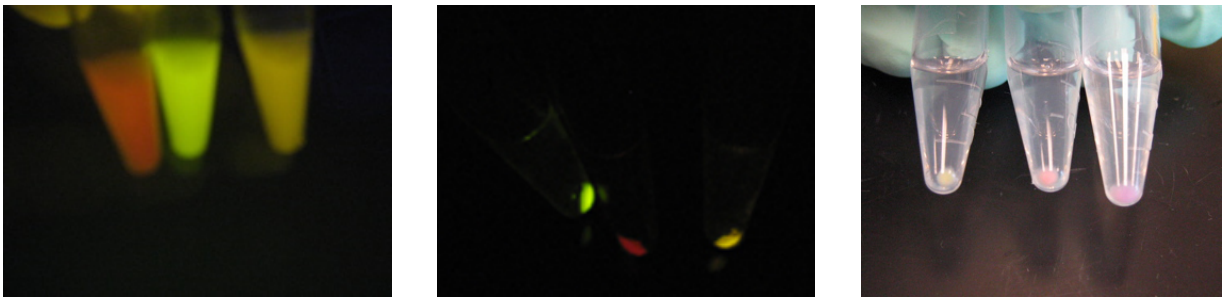
1. Thoroughly mix a fresh, new bottle of Sephacryl S-500, distribute in 10 ml portions, and store in screw cap bottles or centrifuge tubes in the cold room.

2. Prior to use, briefly vortex the matrix and without allowing to settle, add 500 ul of this slurry to a mini-spin column (Millipore) which has been inserted into a 1.5 ml microcentrifuge tube.

3. Following centrifugation at 2K RPM in a table top centrifuge, carefully add 200 ul of 100 mM Tris-HCl (pH 8.0) to the top of the Sephacryl matrix and centrifuge for 2 min. at 2K RPM. Repeat this step twice more. Place the Sephacryl matrix-containing spin column in a new microcentrifuge tube.

4. Then, carefully add 40 ul of nebulized cosmid, plasmid or P1 DNA which has been end repaired to the Sephacryl matrix (saving 2 ul for later agarose gel analysis) and centrifuge at 2K RPM for 5 minutes. Remove the column, save the solution containing the eluted, large DNA fragments (fraction 1). Apply 40 ul of 1xTM buffer and recentrifuge for 2 minutes at 2K RPM to obtain fraction 2 and repeat this 1xTM rinse step twice more to obtain fractions 3 and 4.

5. To check the DNA fragment sizes, load 3-5 ul of each eluant fraction onto a 0.7% agarose



(a) (b) (c)

Figure 10.2: I washed the cells, from an overnight in the BL21(DE3)pLyseE strain with 500 uM IPTG, with Davis minimal media. Image (a) shows proteins mCherry, Citrine, and mOrange (left to right) in suspension. Image (b) shows the same tubes but with the cells pelleted to concentrate the proteins/cells to one location (left to right is Citrine, mCherry, mOrange). Image (c) shows pellets of cultures expressing mCitrine, mOrange, and mCherry in white (ambient) light.

gel that includes as controls, 1-2 ul of a PhiX174-HaeIII digest and 2 ul of unfractionated, nebulized DNA saved from step 4 above.

6. The fractions containing the nebulized DNA in the desired size ranges (typically fractions 1 and 2) are separately phenol extracted and concentrated by ethanol precipitation prior to the kinase reaction.

Bruce A. Roe, Department of Chemistry and Biochemistry, The University of Oklahoma, Norman, Oklahoma 73019 broe@ou.edu

TM buffer shows up as many different things on many different websites, protocol books. The TM recipe that the people which posted the sephacryl recipe have on the website is: 50 mM Tris [pH 8], 15 mM MgCl2.

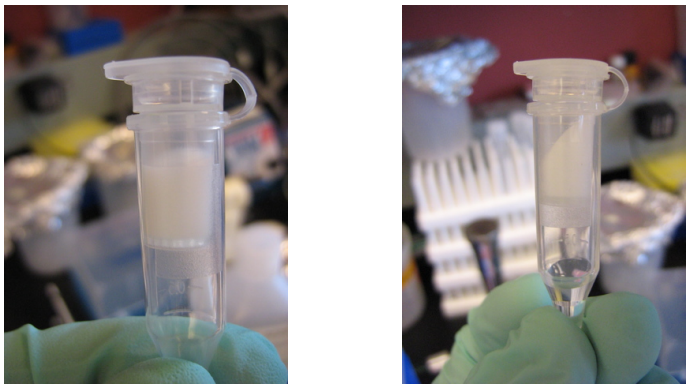
10.2.1 Testing the old protocol

Wed Oct 25, 2006

I want to test the original protocol I got off the web first. I used 2 sources of genomic DNA and a cut plasmid in the size-selection process. The genomics sources were samples 1 and 5 (Figure 5.5) from section 5.2.4 on page 214. I choose the two genomic samples because they provide a nice smear across two different ranges (large for genomic 1 and medium/small for genomic 5 see Figure 5.5). I added the cut plasmid so that I would have a strong reference band at one particular size. Cut pUC19 is around 2700bp.

I tried fractionating each sample individually and all three combined. The all three combined was the only sample that I ran on a gel. It consisted of 500 ng cut pUC19 and 2 μg of each of the genomic samples. The samples were combined into a total volume of 40 μl in TE.

I added the 500 μl of Sephacryl S500 [Amersham] to a Spin-X filter column [Costar] and washed it as described in the section above. Notice that the centrifugation pulls the sephacryl onto a hard slant 10.3, so that at it's thickest the sephacryl extends to the top of the tube, but on the other exterme it is only 0.25 cm or so thick.



(a) (b)

Figure 10.3: 500 μl of Sephacryl S500 (a) was washed 3x with TE. The centrifugal force distributed the sephacryl on a hard angle inside the Spin-X column (b).

I always loaded the samples at the thickest part of the sephacryl. I ran the initial fraction followed by 3 more fractions with 40 μl TM buffer. I ran all 4 appx 40 μl fractions out on a gel along with a non-fractioned sample with the original composition as my fractionated sample 10.4.

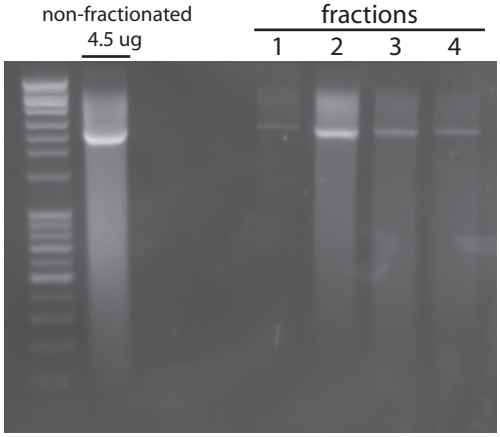


Figure 10.4: 1% agarose gel and 4 sephacryl fractions. an unfractionated sample is included for comparison.

Brief Conclusions: It is vaguely clear that each fraction contains progressively more short fragments and progresively fewer long pieces with each subsequent fraction (Figure 10.4). However, the different between the fractions is pretty miniscule here, and I'm not sure it is useful except to get rid of the smallest of the small fragments. I want to try again with more sephacryl in the hopes that having the minimum thickness of sephacryl increased (see Figure 10.3) will also increase the separation achieved by the fractionation.

10.2.2 Improving the sephacryl protocol

Oct 26, 2006

I want to determine if increasing the amount of sephacryl or changing the MgCl2 concentration (both of which I assume would slow the DNA's migration through the sephacryl) would better the fractionation.

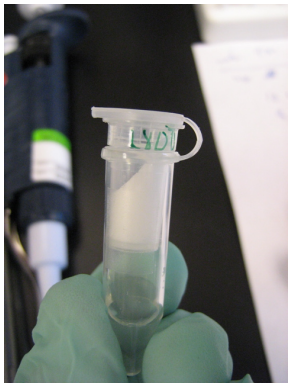


Figure 10.5: Using 850 μl of sephacryl S500 created a much thicker layer at the bottom of the centrifuge created slant.

I used the concentration of all of the DNA components I used last time (genomic samples 1 and 5, plus cut pUC19). This time I ran two samples, one with 750 μl of sephacryl (50% more than last time) and 15 mM MgCl2 (the same as last time) and another with 850 μl of sephacryl (70% more than last time) and 7.5 mM MgCl2 (50% less than last time).

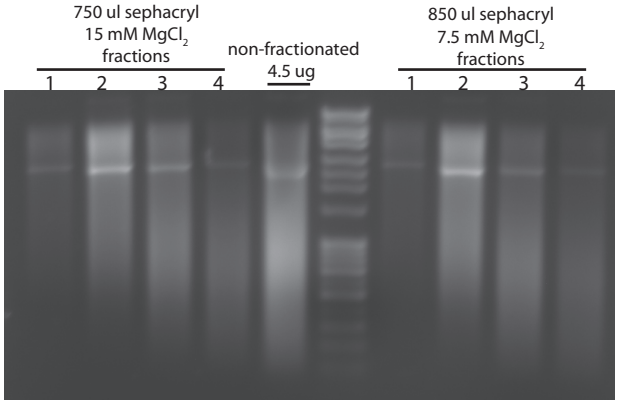


Figure 10.6: 1% agarose gel stained with EtBr. The two different MgCl2 and Sephacryl concentrations are indicated above the gel.

Like last time, I ran each 40 μl fraction in a separate lane on a gel (Figure 10.6). I only ran one non-fractionated sample, since both protocol variants I tried had the same starting material.

Brief Conclusions: The increased sephacryl amount certainly increased the minimum thickness at the bottom of the centrifuge generated slant (compare new Figure 10.5 with the old Figure 10.3). And the resulting size-fractionation was much more marked (Figure 10.6). The 850 μl with the lower MgCl2 concentration seemed to work better. I really would like to try one more time with even more sephacryl, though it's getting to the point that it'll be pretty cramped in there. Perhaps there's something other than Spin-X which will hold more volume at the top (besides the really

long standard chromatography columns which I'd rather not mess with). I should also consider buying the ones from invitrogen and comparing with these home-made ones.

I'd really like to get that initial cut off for the second column to be greater than 500 bp. Right now, it is pretty strong still at 500 bp and weakens to low at around 300 bp and to nothing at around 200 bp.

To Do!!! Try to run with 1 ml or 1.25 ml of sephacrl. Will likely need to spin down 800 μl sephacryl, add 200-400 more, spin down again and THEN wash 3x. Buy fractionation columns from Invitrogen?

10.2.3 Improving the sephacryl protocol, part 2

Thu Nov 9 17:45:53 EST 2006

I'm going to try and fractionate with 1 ml or 1.2 ml of sephacrl. I spun down 800 μl sephacryl, added 200-400 more, and spun down again. The 1 ml column was almost full. The 1.2 ml column was completely full, any more and the lid wouldn't really close properly. Then I washed the column 3x with TE. I still used 2 μg of genomic 1 and 2 μg of genomic 5. However, I didn't have any pUC19. Instead I used 1 μl of cut pNEB193. I'm not sure how much it was? Maybe 200-400 ng total.

One last problem. The original protocol says to spin 2K rpm for 2 minutes between each TE wash. I only used 1 min for the first two. I thought it wouldn't matter, however when I ran the first fraction through and spun for 5 minutes there was way more than my 40 μl starting material (probably 60-80 μl).

I ran the fractions on a 80 ml, 1.5mm comb, 1% agarose gel for 50 minutes at 90V [3] (Figure 10.7).

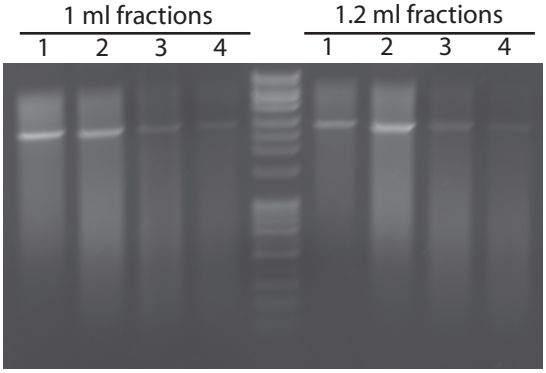


Figure 10.7: 1.0% agarose gel of sheared genomic DNA and plasmid run fractionated through columns with 1 ml and 1.2 ml sephacryl S500.

Brief Conclusions: Fri Nov 10 10:42:09 EST 2006

Next time during the TE wash steps, I need to make sure to centrifuge for 2 minutes. Perhaps use 3 or 4minutes for the last wash centrifugation. I'm still not entirely pleased with the size separation

[3] 80 ml was just enough space for the entire sample

this gives me (Figure 10.7), but I'm kinda running outta things to try. It might be good enough to just use the first two fractions. I just hope the cloning efficiency of the 300mers doesn't overwhelm the more abundant longer pieces. Another thing to consider is using the invitrogen premade columns. I should buy some and give them a try. But their expense make them unsuitable for everyday use.

To Do!!! buy invitrogen sephacryl columns; test the ability of the microcon columns to remove *large* numbers of short adaptors and blunt ligated adaptors.

10.3 Comparison of short DNA fragment removal with Microcon 30, 50, or Qiagen PCR purification kit

Whenever I use adaptors in a ligation reaction, the huge excess necessary to prevent incorrect blunt ligations makes it dang hard to run on a gel. For an agarose gel you get a giant band where the primers are, for a polyacrylamide gel you get a giant black spot and a huge amount of DNA noise that pretty much makes the rest of the gel unreadable. How can I get rid of these pieces? When adaptoring cDNA I used a Qiagen PCR column to remove them, but there was still a lot left (cite figure). I'm thinking maybe one of the microcon super-small filter devices from millipore might provide better exclusion. Or perhaps some combination of the two. For sure one nice feature of the Qiagen column is that it does effectively remove very large fragments (see the cDNA cloning chapter).

Ilaria gave me primers to amplify at 80mer and 120mer fragment of pLtet. These sizes are right at the boundary of what the different filtering and concentrating can keep/reject. My hope is to keep these bands strong while virtually eliminating the shorter primer bands.

10.3.1 preparing the 80mer and 120mer PCR products

Nov 15, 2006

I used Phusion Taq [Fermentas] master mix to produce blunt-end products. The reaction was 2 μl 10 μM primer (400 nM), 1 μl plasmid (11 ng), 22 μl H_2O , 25 μl Phusion master mix. I ran two reactions for the 80mer and two reactions for the 120mer. The two reactions for each were combined and cleaned up using a Qiagen PCR purification kit. Yields were:

Sample	DNA (ng/ul)	260/280	260/230	total yield
80 mer pLtetO	55.0	1.80	2.11	1.65 μg
120 mer pLtetO	45.0	1.77	2.33	1.35 μg

Brief Conclusions: Minus the wierd upside-down cross thing, so far so good.

10.3.2 preparing the test DNA

Nov 16, 2006

I'm going to use a similar composition to what I used with the sephacryl but with the addition of some shorter stuff. Per rxn I'm using 2 μg genomic 1, 2 μg genomic 5, 200 ng 80 mer, 200 ng

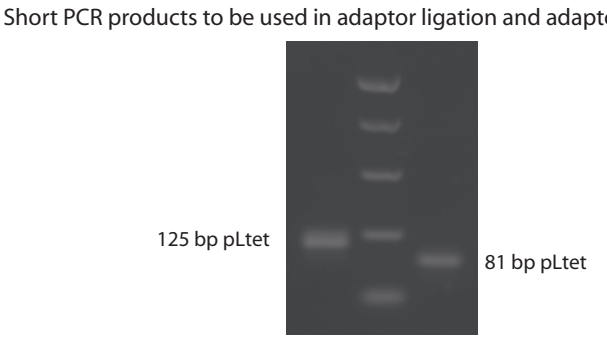


Figure 10.8: 2.0% agarose gel of Phusion PCR amplification of pLtet. Looks like an upside-down cross - yikes!

120 mer (appx 4 μl of each), and 25 fold excess of each small primer pair (see section 6.7.1 page 279), which corresponds to 1.8 μg of the 30mer and 900 ng of the 15mer and 18 μl from the 10 μM primer stock. In total, this is around 7.1 μg of DNA.

I made 4x this amount to try with a Qiagen cleanup, a YM30, YM50, and no column. I need to make sure to run the no column lane far from the others to prevent saturation problems.

Fri Nov 17 15:16:43 EST 2006

The Qiagen PCR prep DNA was eluted in 30 μl . The microcon were spun at 14,000g for 12 minutes and then eluted. The elution was too small so I added around 30 μl to the top of each and weakly vortexed for 30 seconds.

The yields for all of the different methods were:

Sample	volume	DNA (ng/ul)	260/280	260/230	total yield
non-processed	80 μl	99.0	1.93	0.96	7.92 μg
microcon YM30	17 μl	338.0	1.85	1.72	5.75 μg
microcon YM50	15 μl	376.9	1.83	1.70	5.65 μg
Qiagen PCR	30 μl	104	1.97	1.89	3.12 μg

I ran 1 μg of each sample in a total volume of 20 μl (including dye) on a 15% TBE Urea gel for 50 minutes at 190V. The gel was stained 20 minutes in EtBr and destained 10 minutes in H_2O (Figure 10.9).

Brief Conclusions: Fri Nov 17 15:16:43 EST 2006

Well, these things don't do a terribly good job of removing what they said they remove. The microcon columns to a great job of concentrating and not lossing much DNA (at least relative to the Qiagen column), but the primers weren't really removed (Figure 10.9).

10.4 Comparison of short DNA fragment removal with Qiagen PCR purification kit and ChargeSwitch with 1x, 2x, 3x, 4x N5 buffer

Mon May 14, 2007

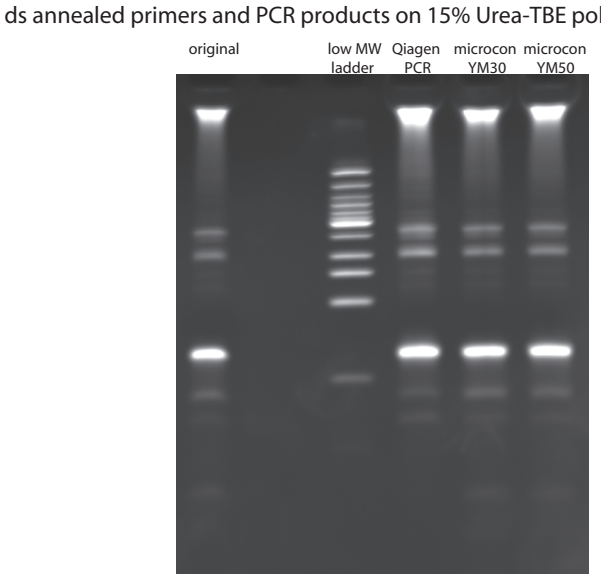


Figure 10.9: 15% TBE Urea gel

Once again, I'm continuing my thus far unsatisfactory search for a non gel-purification based method to remove short DNA fragments. Removing the short DNA pieces without the inefficiencies of gel-purification would really speed things up and improve my ability to do several of the techniques I'm working on.

This time I'm comparing my default Qiagen PCR purification with a newer magnetic bead based method call ChargeSwitch from Invitrogen. I says in the Invitrogen manual that it works for purification of 25 μl - 50 μl PCR reactions. It also mentions that by altering the concentration of the N5 buffer, you could change the minimum DNA size retained by the purification kit (higher N5 concentration = fewer short pieces).

I tested their statement with the 50 bp ladder from NEB, so that I could look for DNA retainment at a varieaty of sizes. I made 25 μl samples (2 μl NEB 50 bp ladder and 23 μl TE). I tested the Qiagen kit and the ChargeSwitch kit with 1x, 2x, 3x, and 4x N5 buffer. The DNA was then run on a 2% agarose gel 10.10.

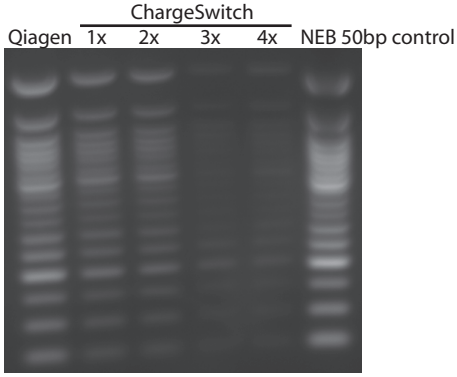


Figure 10.10: 2% agarose gel showing short DNA size removal (or rather the lack thereof) of Qiagen PCR purification kit and ChargeSwitch with different concentrations of N5 buffer.

Versadoc calibrated bands from Figure 10.9							
Lane	**Base Pairs**	**Peak Int**	**Average Int**	**Trace (Int x mm)**	**Relative Qty**	**Calibrated (ng)**	**Normalized (Qty)**
1	163	1478	1006	2717	5	38	
1	102	1531	1055	2742	5	38	
1	31	4095	2052	9027	16	206	
1	22	1074	889	2134	4	41	
4	178	2003	1314	4337	7	73	
4	108	2239	1410	3667	6	59	
4	32	4095	1891	10969	17	261	
4	22	834	751	2402	4	34	
5	181	1562	1069	2886	5	40	
5	110	1671	1119	2798	5	39	
5	32	4095	2083	9583	16	221	
5	22	1103	932	2330	4	33	
6	184	1616	1097	3073	5	42	
6	113	1691	1160	2900	5	40	
6	32	4095	2070	10352	16	243	
6	22	1024	902	2165	3	40	

Brief Conclusions: To really say something more precise about this result I could use the versadoc to estimate the concentration of the bands of interest (in particular 50bp and 100bp). But generally, this looks like another size-selection disappointment. The ChargeSwitch yield is much worse than the Qiagen, and the removal of even the 50bp piece by any of the kits is negligable. Also notice that the main influence of increased N5 concentration is decreased yield.

10.5 Comparison of Qiagen Column Gel cleanup and QiaexII gel cleanup

Thur Jul 5, 2007

I think there is a need to really understand these kits that we use all of the time in the lab. We need an understanding that goes beyond just the printed stuff that comes with the kit. I want to be able to gel purify stuff, obtaining the highest yield with the cleanest DNA.

I'm testing the two kits from Qiagen: the column based method and the QiaexII bead based method. To test these methods I ran 1x NEB 2-log ladder and 0.5x NEB ladder on a 1% TAE agarose gel for 45 minutes. For each ladder, I cut and purified 2 bands (one big and one little) for each kit (so 4 total for each kit). I also cut two bands from lanes with no DNA as a negative control. I purified them according to each kit's instructions. I used 550 μl of buffer QG in all of the column methods to solubilize the gel. And I used 550 μl of buffer QX1 in all Qiaex methods to solubilize the gel. I eluted all of the purifications into 30 μl EB buffer. To quantify the yield of each purification, I used 20 μl of the 30 μl elution with the HS dsDNA Qubit fluorescence quantification platform.

The results were:

kit	ladder	band (bp)	starting DNA (ng)	cleaned up DNA (ng)	recovered
column	none	blank	0	0.543	-
column	0.5x	100	30.5	16.5	0.54
column	0.5x	500	62	39.6	0.64
column	1x	200	32	11.79	0.37
column	1x	3000	120	69.0	0.58
qiaexII	none	blank	0	0.759	-
qiaexII	1x	100	61	24.87	0.41
qiaexII	1x	500	124	78.6	0.63
qiaexII	0.5x	200	16	9.12	0.57
qiaexII	0.5x	3000	60	26.79	0.45
mean recovery from Qiagen column					0.53
mean recovery from QiaexII					0.51

raw qubit readings for this data

Brief Conclusions: Sat Jul 14 19:44:55 EDT 2007

Both of these Qiagen kits performed quite similarly. And to my eye, I don't notice any great differences in recovery via gel extraction with differing sizes and amounts of DNA. I'm not too surprised by the average recovery of 50%; it's actually a little higher than I expected. But these recoveries are much lower than the numbers posted on Qiagen's websites. For the column kit, Qiagen claims up to 95% recovery (I only got up to 64% recovery) with a typical recovery being around 80%. For their Qiaex II kit, they claim a recoveries of 60-95%. Maybe the difference is that they started with *way more* DNA than I did for each size. They used 2 μg vs my 16-124 ng of DNA. Perhaps, I'll try again some day with more DNA, but it's not very common that I have 2 μg of a single size of DNA that I want to purify on a gel – that's too much for my everyday needs.

10.6 Comparison of short DNA fragment removal with PCR Purelink kit from Invitrogen

Sat Jan 19, 2008

I still haven't found a non-gel based way to remove short DNA efficiently. I noticed in the Invitrogen Molecular Biology catalogue that they have a PCR purification kit that comes with a special buffer for removing dsDNA less than 300bp. So I bought the kit and decided to try yet another method to remove short DNA fragments.

I ran 5 μg of 50 bp ladder [NEB] through:

1. Purelink PCR column with standard buffer
2. Purelink PCR column with HC buffer
3. Qiagen PCR purification column

I eluted all three samples into 50 μl of the elution buffers that came with the respective kits. The yields were:

Sample	DNA (ng/ul)	260/280	260/230	total yield	loss
1	61.2	1.90	2.16	3.06 μg	38.8%
2	70.1	1.92	2.19	3.51 μg	29.9%
3	82.4	1.85	2.14	4.12 μg	17.6%

I ran 1 μg of each sample on a TAE agarose gel. I also ran 1 μg of the ladder that had not been run through the column (Figure 10.11).

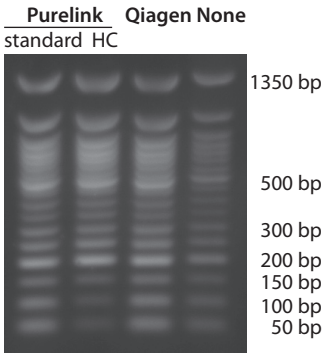


Figure 10.11: It appears that the HC buffer did remove a part (but certainly not all) of the DNA below 300bp.

Brief Conclusions: The HC buffer with Purelink certainly did a better job than the Qiagen column at removing the short fragments. Although the manual stated 300 bp, the real cutoff seems to be more like 200 bp with it being more efficient at removal the smaller the fragment is. I'd like to try again, but this time to run the wash buffer and/or the binding buffer across the column multiple times to see if I can wash the shorter fragments through a little better. As it is, it appears that this kit might to a better job of removing the excessive amount of adaptors (compared with the Qiagen column) that I have after adaptoring my cDNA.

10.7 diffusion of DNA when loading agarose gels

Tues Jul 10, 2007

When I load agarose gels, particularly with DNA that has been cleaned with a Qiagen PCR purification kit, every once-in-a-while it happens that the stupid sample diffuses really fast right after I pipette it into the well of the gel. Needless-to-say this results in a faint band with no chance for quantitive comparisons across samples, but why does this happen sometimes? It seems to happen more with TAE than TBE, but I don't have data to confirm this hunch. My guesses as to whats happening are: 1) eluting into EB buffer contains no salt so EB DNA diffuses away; 2) EB contains no EDTA; 3) trace EtOH from the purification is causing the diffusion.

To test this, I took 1.5 μl of 25 bp ladder [invitrogen] added it into a 10 μl total volume of EB, EB + 15% EtOH, TE, STE. I loaded each sample onto a 2% TAE EtBr agarose gel (Figure 10.12). When I loaded the samples, they all seemed to sink pretty well – no fast diffusers. Which kinda sucked, because it did really relect what I figured would happen, which is that some of the samples would diffuse out quickly and others would not.

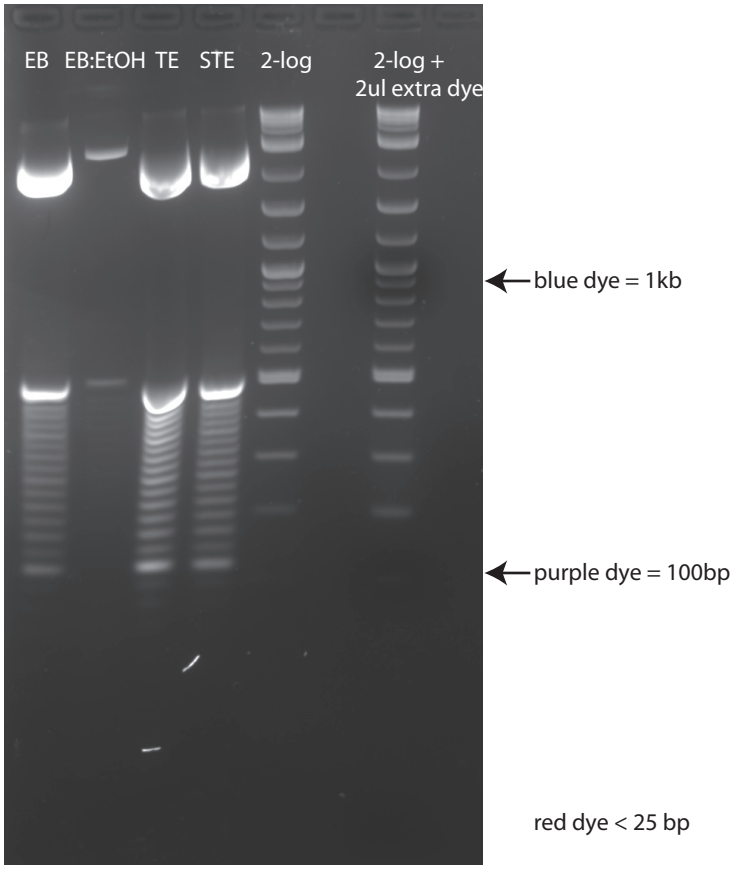


Figure 10.12: 2% agarose gel run at 110 V for 80 minutes

Brief Conclusions: Err, make sure you don't have too much EtOH in your sample. Clearly this makes your sample disappear. As far as the other three go, it is less clear. It seems that EB is a little fainter and fuzzier then the other 2 samples. I'm not sure if the smeary TE lane was a fluke or not, but the 2-log ladders were both in TE and they ran fine. I don't know if the DNA will come off the PCR purification column with a salty buffer, since a high salt buffer is used to bind the DNA in the first place (not sure if it is the pH the salt or both that matter for the binding). But it would be easy to elute into TE and add the appropriate amount of salt to bring the NaCl concentration to 50 mM prior to running the gel.

10.8 Comparison of DNA loss with Qiagen DNA purification vs Qiagen MinElute DNA purification

Fri Feb 15, 2008

It would be convenient to have a smaller volume of DNA sometimes (particularly to run on thin gels). I typically use the Qiagen PCR purification columns to clean up my DNA reactions. I want to compare that kit with the Qiagen MinElute kits, which allow you to elute in down to 10 μl of buffer (the standard PCR purification minimum is 30 μl).

To test each kit, I prepared a mastermix of 50 bp ladder [NEB] and aliquoted around 4 μg of the ladder in 30 μl of buffer into 6 tubes. I ran 3 standard Qiagen PCR cleanups and 3 MinElute ones.

I eluted the standard cleanups into 30 μl and the MinElutes into 10 μl . To prevent any calculation biases with the Nanodrop spec that I used to quantify the results, I added 20 μl of buffer to each of the MinElute cleanup elutions, so that the MinElutes and the standard cleanups would all be in 30 μl of buffer. The specs were:

Sample	**DNA (ng/ul)**	**260/280**	**260/230**
prior to cleanup A	115.1	1.85	2.11
prior to cleanup B	115.2	1.86	2.11
standard A	103.1	1.78	2.10
standard B	109.4	1.82	2.09
standard C	103.1	1.80	1.85
minElute A	88.5	1.80	2.10
minElute B	101.7	1.80	2.05
minElute C	91.5	1.82	2.09

The mean concentration of the standard cleanup samples was 105.2 ng/μl , which represents a loss of 8.6%; The mean concentration of the MinElute cleanup samples was 93.9 ng/μl which represents a loss of 18.4%.

Brief Conclusions: The Qiagen manuals were right on the money in saying that a standard kit gives less than a 10% loss and a minElute kit gives less than a 20% loss.

Appendix A

Techniques

A.1 Nucleic Acid Determination by Spectrophotometry

- 260 absorbance positively indicates the nucleic acid quantity
- 260/280 absorbance ratio positively indicates protein contamination (best range is between 1.8 and 2.0)
- 230/280 absorbance ratio positively indicates protein, solvent, and salt contamination (best range is > 2.0)

Values found outside these acceptable ranges are colored red in the text of this notebook.

A.2 Pipetting

A.2.1 Multichannel pipetting

Multichannel pipetting brings a little sanity to the montonous life of a middle throughput experimentalist. Here are a few tips that I've figured out or picked up from other people along the way.

Are you having problems sucking the same amount of sample from each channel?

I think this is a never ending problem and the most frequent one. But it doesn't have to be there if you do a little work. I assume your pipettor is calibrated, so the error is your own fault not the instruments. If it's not calibrated once a year, it should be.

When I first started I had a tendency to just pipette anyways if the differences were small between because I thought they would average out in the end and I was lazy. This is a bad approach. It is usually the same channel having the same error because there isn't a good seal, and the seal often gets worse all of this could lead to some nice correlations in your figures that you spend years exploring as scientifically interesting results that can only repeated with your pipettor.

The easiest way to get a big boost in accuracy is to buy the tips recommended by the manufacturer. These are usually the most expensive tips available for your pipettor, and they are typically made

by the manufacturer of your pipettor. The second easiest way, but one the requires a little work once a week or so is silicon grease. Adding a little (and I stress *little*) amount of this magic goopy stuff to each channel allows the channels to form a much better seal with the tip, so when you set the pipetter to suck up 10 μl you suck 10 μl with each channel rather than a random number between 5 and 10 μl in each and the rest air.

This trick was taught to me by Jamey Wierzbowski. I'm not sure how he does it but here's how I do it:

1. dip a Kim wipe into a container of silicon grease trying to get a small amount on it
2. rub the kim wipe / grease across each side of the channels coating them well but not globbing it on there; avoid getting it onto the holes where the air is aspirated from
3. use your fingers (with a glove on of course) to rub the grease already on the channels into those hard to reach places between the channels (if you have giant fingers you should probably figure out a different way to do this
4. wipe of the channels with kim wipes until you can no longer see any grease on the channels
5. pipette some water with all 12 channels and if when you eject the tips they don't shoot off or shoot off very hesitantly, you have way too much grease on there. wipe more off with kim wipes and be through this time you lazy bum.
6. You are now ready for pipetting heaven. Pipette and enjoy seeing exactly the same amount of liquid in each tip.

Valuable Lesson: Rub a little silicon grease onto your multipipettor channels regularly to keep them pipetting accurately.

Wed Feb 15 11:16:18 EST 2006

On the last 384-well plate of my 9000 qPCR reactions in the initial ChIP-Chip study, I made a discovery that is more important than the silicon grease. The most important factor is to have the tips made by the same manufacturer as the pipettor. I was using generic tips from Fisher that were supposed to fit the Finnipipette multichannel we have. They did kinda, but it is no comparison to the real Finntips. The tips are more expensive then the generic ones however, by using refill pages the price is almost exactly the same (one-hundredth of a penny more per tip) and I can save the environment (FINNTIP 10 REFILL KIT Fisher Part # 14-245-149, FINNTIP 10 REFILL Starter 14-245-148). With the starter you get 4 tip holders, so it is probably necessary to buy 2-3 of those.

Valuable Lesson: Use tips made by the manufacturer of your pipettor – at least for multichannel pipetting

Appendix B

Recipes

B.1 Antibiotic stocks

B.1.1 Norfloxacin / Ciprofloxacin

Reagent	Amount per 200 ml
5 N NaOH	400 μl
nor or cipro	0.25 g

This stuff is hard to get into solution. The NaOH is supposed to help. Stock is good at least 1 month in at 4°C. Freezing for longer storage seems to be ok (according to Jamey; I haven't tried yet). Takes *very* little of this stuff to tear up some bacteria. MIC is around 150 ng/ml. That is only 6 μl of stock solution in 50 ml. I use 75 ng /ml to cause DNA damage but not kill them (does slow them down a little).

B.2 Common Buffers

B.2.1 Phosphate Buffered Saline (PBS)

Reagent	Amount per Liter
KCl	0.2g
NaCl	8g
KH_2PO_4	0.240g
Na_2HPO_4	1.44g

Adjust pH to 7.4 with HCl, I always check with the pH meter and it never has needed adjusting. Autoclave to sterilize.

B.3 Agarose gel buffers what?

B.3.1 Tris-Acetate-EDTA (TAE)

I use the 4 L 50X TAE dispensor from Fisher.

Use 5-8 V/cm (higher V for short fragments, lower V for long ones)

B.3.2 Tris-Borate-EDTA (TBE)

I use the 4 L 10X TAE dispensor from Fisher.

Use 5-35 V/cm (higher V for short fragments, lower V for long ones)

B.3.3 Sodium Boric Acid (SB)

This is a buffer for running very fast gels (10x faster than TAE) at high voltage. Works best for shorter fragments (less than 2kb). Doesn't work as well with samples containing higher salt concentrations (e.g. restriction digests). More info can be found in Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis..

Reagent	Amount per Liter
NaOH	8 g
Boric Acid	45 g

Fill to 1 L with H_2O .

Use 5-10 V/cm (higher V for short fragments, lower V for long ones)

B.4 DNA and RNA storage buffers

B.4.1 Annealing of two oligos

Use STE Buffer (10 mM Tris pH 8, 50 mM NaCl, 1 mM EDTA). The salt allows the oligos to anneal.

B.4.2 Resuspension of lypholized primers

I follow the FAQ from IDT: Dissolve the stock oligo in TE. Make the freezer stock at 100uM by adding a volume of TE ten times the number of nanomoles of DNA present in the tube (it is noted on the spec sheet provided by IDT.

For jobs requiring 24 or more primer pairs I order them in plates frozen in TE at 100uM concentration. I put the pairs on consecutive rows to ease the construction of primer pairs.

I typically aliquot to a stock at a concentration of 2uM. And use 1.5μl in a 20μl PCR.

B.5 Midiprep Solutions

Added: Sun May 28 18:41:34 EDT 2006. Last Update: Sun May 28 18:41:34 EDT 2006.

B.5.1 Alkaline Lysis Solution 1

Just a little something to resuspend the guys in. I add the RNAse Cocktail fresh right before using. 1 μl of RNAse cocktail per 200 μl of Alkaline Lysis Solution 1. Store at 4°C.

Reagent	Amount	Stock Conc.	Amount for 50ml
Sucrose	50 mM	-	0.45 g
Tris-HCl [pH 8.0]	25 mM Tris	1M	1250 μl
EDTA	10mM	0.5M	1 ml

B.5.2 Alkaline Lysis Solution 2

This stuff lyses the cells.

Reagent	Amount	Stock Conc.	Amount for 1 ml
0.2 N NaOH	0.2 N	1 N	200 μl
SDS	1%	20%	50 μl

B.5.3 Alkaline Lysis Solution 3

This neutralizes the lysate.

Reagent	Amount	Stock Conc.	Amount for 50 ml
potassium acetate	3 M	5 M	30 ml
glacial acetic acid	-	-	5.75 ml

B.6 ChIP Recipes

B.6.1 Palsson lysis buffer

Palsson lab adapted from Grossman protocol.

Reagent	Amount	Stock Conc.	Amount for 50ml
Tris-HCl	10mM Tris	1M	500 μl
NaCl	50mM	5M	500 μl
EDTA	10mM	0.5M	1 ml
Sucrose	20%	-	10 g
Ready-lyse lysozyme	0.2μl per 500 add fresh	-	

B.6.2 Palsson 2x IP buffer

Palsson lab adapted from Grossman protocol.

Reagent	Amount	Stock Conc.	Amount for 50ml
Tris-HCl	200mM Tris	1M	10ml
NaCl	600mM	5M	6ml
Triton X-100	4%	-	2ml
1mM PMSF	added just prior to use	100mM	500 μl

B.6.3 Dilution Buffer

Taken from www.abcam.com

Reagent	Amount	Stock Conc.	Amount for 50ml
Triton X-100	1%	-	500 μl
EDTA	2 mM	0.5 M	200 μl
NaCl	150 mM	5 M	1.5 ml
Tris-HCl pH 8	20 mM	1 M	1 ml
PMSF	1 mM added just prior to use	100 mM	500 μl

B.6.4 Low Salt Immune Complex Wash Buffer

Reagent	Amount	Stock Conc.	Amount for 50ml
SDS	0.1%	-	250 μl (of 20%)
Triton X-100	1%	-	500 μl
EDTA	2 mM	0.5 M	200 μl
Tris-HCl pH 8	20 mM	1 M	1 ml
NaCl	150 mM	5 M	1.5 ml

B.6.5 High Salt Immune Complex Wash Buffer

Reagent	Amount	Stock Conc.	Amount for 50ml
SDS	0.1%	-	250 μl (of 20%)
Triton X-100	1%	-	500 μl
EDTA	2 mM	0.5 M	200 μl
Tris-HCl pH 8	20 mM	1 M	1 ml
NaCl	500 mM	5 M	5 ml

B.6.6 LiCl Immune Complex Wash Buffer

Reagent	Amount	Stock Conc.	Amount for 50ml
LiCi	0.25 M	5 M	2.5 ml
NP40	1%	-	500 μl
deoxycholate	1%	solid powder	500 mg
EDTA	1 mM	0.5 M	100 μl
Tris-HCl pH 8	10 mM	1 M	500 μl

Appendix C

Protocols

C.1 Reverse transcription

C.1.1 Superscript II/III

C.1.2 How much?

Use 200 U for the first 1 μg , then use 200 U for each additional μg .

C.1.3 Heat stability

To Do!!! Heat superscript II and III to 95C for N minutes (try 2, 5). Cool and try reverse transcription. See if either still works.

C.2 Cloning

C.2.1 Buying chemically competent cells

For routine cloning, I use the DH5α Subcloning competent cells [Invitrogen 18265-017]. They don't recommend these for cDNA cloning. For cDNA cloning they recommend using MAX Efficiency DH5α Competent Cells or better. They cost 2x as much. The subcloning cells are so cheap, I hesistate to ever make competent cells for cloning again, since theirs are better than mine. I'll use my own cells only when I need to clone into a wierd strain.

C.2.2 Cloning vector into purchased competent cells

This protocol is for when you have intact (non-ligated) DNA. It is designed for speed not efficiency.

1. thaw a 50 μl aliquot of competent cells on ice
2. add 250 pg - 5 ng of plasmid
3. flick tube to mix plasmid with vector

4. leave on ice 15 minutes (I should try making this time even smaller)
5. heat shock 20 seconds at 37 C
6. place tubes back on ice for 2 minutes; during the 2 minutes add 350 μl of SOC (or LB) to the tube
7. incubate at 37 C with shaking for 45 min
8. spread 30 μl onto an agar plate with proper antibiotic resistance

C.2.3 Cloning vector into purchased competent cells

This protocol is for tranforming ligated DNA. It is designed for efficiency not speed. The protocol is similar to the Invitrogen protocol that comes with the cells, except the dilution with SOC is much less.

Use 50 ng of vector, 1 μl of Ligase, and a 3:1 insert to vector ratio in a 20 μl reaction volume.

1. thaw a 50 μl aliquot of competent cells on ice
2. add 2 μl of the ligation mixture
3. flick the tube a few times to gently mix the cells / ligated plasmid
4. incubate tubes on ice for 30 minutes (Ilaria uses just 15 minutes)
5. heat shock for 20 seconds in a 42°water bath
6. place tubes on ice for 2 minutes
7. add 250 μl of SOC
8. incubate the cells at 37°C for 1 hr at 225 rpm (Ilaria grows 45 min at 300 rpm)
9. plate 100-150 μl on a pre-warmed agar plate[1] with appropriate antibiotics
10. wait 12-24 hrs (typically overnite) for colonies to grow

Bio-cheats: Don't throw away the cells you leave on the bench. Leave them on the benchtop. If the next day you don't have colonies from your transformation, carefully suck of all but 70-100 μl from the transformation. The cells will have settled to the bottom of the tube overnite, so don't disturb the pellet before you aspirate. Now flick the tube to resuspend the cells in the smaller volume. Plate the entire remaining volume onto a new prewarmed agar plate with the appropriate antibiotics.

[1] if doing blue-white screening, add 40 μl of *Blue-White Select Screening Reagent* [Sigma] 10 min before plating the cells

C.2.4 Chemically competent cell preparation

C.3 ChIP Protocols

This protocol is a mixture of three protocols from abcam, upstate, upstate Tips/Protocol, and palsson et al 2005.

C.3.1 Lyse, Crosslink, Shear

The following protocol should be started in the morning and will finish later on in the evening (allocate 10hrs). The morning will be only sporatic work but as the day goes on and on, full time must be devoted to the experiment (from the lysis step forward).

Growth

1. Grow stain(s) in 50ml of liquid media in a 250ml flask with appropriate antibiotics
2. add IPTG or other inducer as needed for induction of any cloned constructs (if you have a specific non-cloned gene or tagged gene in the genome, this step is not necessary)
3. take M, Nml samples as replicates from the same culture (replicates in different cultures are important too; typically I use M=1 and N=15). I put this in a 15ml corning centrifuge tube.

Crosslinking

1. add 37% formaldehyde to 1% total concentration for each (400μl for 15ml sample) and mix by inversion (invert tube 10 times). Incubate 10 min at room temp. Do NOT crosslink too long!
2. pellet cells by centrifugation for 10 min at 3200g (this doesn't pellet ALL of them but enough for downstream purposes; if OD is low (< 0.3) all of sample will pellet)
3. decant supernant in formaldehyde/media hazardous waste bin
4. wash cells 2x in ice-cold PBS (spin 3200g for 8 min) (from here on, keep samples on ice to prevent proteases from messing up the experiment; I use 2/3 initial volume for washing: e.g. with 15ml of cells I use 10ml of PBS to wash)
5. prior to lysis remove any remaining drops of PBS that wouldn't go away by decanting with a P1000 pipettor

Lysis

1. add 0.4μl Ready-Lyse per ml of lysis buffer into a lysis buffer master mix (just before using)
2. begin lysis by adding 500 μl ChIP lysis buffer and incubate for 37C for 30 min (no shaking)
3. add 10μl of 100mM PMSF and 2μl RNAse cocktail per ml of 2x Palsson buffer (just before using)
4. add 500 μl 2x Pallson IP buffer with 1 μl fresh RNAse cocktail [Ambion] and incubate at 37C with shaking at 300rpm

5. lysate should be clear like water (if it's not, I don't know maybe you did something wrong. For me it's always looked like slightly-soapy water)

6. transfer the 1ml lysate to a 1.5ml eppy tube with a P1000

Shearing

1. sonicate samples using a Branson 250 Sonifier for 30 secs at 20% power (more consistent results can be obtained by using a digital Branson where the exact time can be set. also, it helps to rig up the stand holding the sonifier tip to also hold the eppy tube so the tip stays in the same part of every eppy tube sample.

2. keep samples on ice for at least 1 min between sonifications

3. sonicate each sample 4 times for shearing range of between 1000bp and 100bp with an average around 500bp (See Figure 1.24 for an example)

12 samples takes around an hour; dreadfully slow and boring; I play my ipod under the protective sonicator earphones. After half-an-hour this becomes painful if you have the ipod earbuds, so adjust them until it doesn't hurt anymore. Shearing is much less tedious with some tunes.

Quantification Part 1

1. remove 100 μl from each sample to quantify the amount of DNA in each lysate and to verify the shearing range

2. place the remaining 900 μl in the -80C freezer for use in immunoprecipitation (it'll be more than enough)

3. add 350 μl of water to the 100 μl sample (450 μl total volume); add 5 μl of 20mg/ml proteinase K added;

4. reverse crosslinks overnite in a 65C heat block

C.3.2 Sheared DNA yield / Beginning of immunoprecipation

This protocol also takes a day but with lots of breaks.

Quantification Part 2

Although the DNA yield from this step is pretty high, I can never see a DNA pellet (even though when doing precipitations with much smaller amounts I have seen one; perhaps cause it's sheared?). This used to worry me; now I just trust that it'll work and make sure to suck out the liquid on the side of the eppy tube where the DNA shouldn't be when doing my ethanol precipitation.

1. separate DNA from proteins by phenol:chloroform extraction using gel phase lock tubes; I use the gel-phase tube only for the second half (the chloroform only step) and I get cleaner DNA than when I use the gel phase lock for both steps. Plus, it seems to smell less like phenol this way.

2. add 1/10th volume NaAcetate and 2 μl of N mg/ml glycogen as a DNA carrier (not sure this does anything at all, but doesn't hurt)

3. add 1ml ethanol
4. place in -80C for 30min
5. spin at 0C at maximum rpm for 20min
6. remove supernant with vacuum at 50 mbar, place tip on/near side opposite the g-force (i.e. the side the DNA is not stuck to)
7. add 1ml of cold 70% ethanol (some people say to resuspend DNA pellet here by vortexing, I don't).
8. spin at 0C for 5 min
9. remove supernant with vacuum at 50 mbar, place tip on/near side opposite the g-force (i.e. the side the DNA is not stuck to)
10. air dry in fume-hood to allow ethanol to evaporate
11. resuspend DNA in 100 μl of TE
12. quantify using spectrophotometer (e.g. a Nanodrop)

The cleaned up DNA should be frozen at -20C and saved, as it will serve as the positive control for all downstream qPCR reactions. 500-600ng of each sheared DNA sample should be run an a 1.5% agarose gel to verify shearing range. Examples of yields vs OD can be seen on pages 95 and 102. For the next step it is helpful to have the DNA concentration over 250ng/μl .

Immunoprecipitation

Immunoprecipitations are begun with equal starting DNA (25 μg). So DNA is first quantified using the protocol above. The following steps I do in 2ml eppy tubes.

1. divide the 900 μl left of the sheared lysate into N replicates where N is divisible by 3. Each sample should contain 25 μg of DNA. Label these **A**, **B**, and **C** (I normally only use N=1).
2. dilute all samples 1:10 in dilution buffer (if your DNA isn't concentrated, dilute as much as possible, should still work).
3. preclear solution by adding 40μl of agarose beads and rotate at 4C for 90 minutes
4. spin at 1000g for 1 min and transfer supernant to a new tube
5. add 2ug of the correct antibody to sample **A** (e.g. if your tag is myc put anti-myc here)
6. add 2ug of an incorrect antibody to **B** (a negative control) (e.g. if your tag is myc put anti-Xpress here)
7. **C** is a negative control with no antibody
8. rotate all samples overnite at 4C, preferably in 2ml eppy tubes (these allow better mixing during rotation because the liquid doesn't get stuck as poorly in the bottom; they do however make the later bead washing more difficult because it is hard to see the beads in the 2ml tube)

By pulling all three of these from the same lysate you get nice samples for detecting the different between A, B, and C. Independent sample replicates are important as well. Just try to start with the same concentration of DNA in every immunoprecipitation.

Last, I'm not convinced the **C** negative control is useful. In the end, I might dump this control as it is expensive to have three controls (one positive and two negatives), takes up valuable space on the qPCR plate, and adds only a small sanity check.

C.3.3 Immunoprecipitate Bead washing

This protocol takes 3-5hrs. Be careful not to suck out the beads by mistake. I use a vacuum with a vacuum regulator adjusted to make the suction very weak (I use approx 50mbar). I'm also conservative with my washing steps and leave 100 μl or so in the tube each step except the final elution step which I work hard to remove all liquids and no beads before starting.

Centrifuge at 4C for 1 min at 1000g; be careful not to disturb the bead pellet when moving samples. Keep samples on ice, unless specified otherwise. Rotate each wash 5 min in cold room, except for TE wash, which is rotated at room temp.

Add beads add 60μl of salmon sperm / agarose beads (I buy the mixture from upstate). Rotate at 4C for 2hr.

Bead Washing

1. wash beads 1x with low salt wash
2. wash beads 1x with high salt wash
3. wash beads 1x with LiCl wash
4. wash beads 2x with TE (steps at room temp from now on)
5. elute by adding 225 μl fresh elution buffer and rotating for 15min. Keep the supernant in a 1.5ml eppy tube.
6. repeat elution with 225 μl more elution buffer and combine supernant with previous supernant.

Crosslink reversal

1. add 10 μl 5M NaCl solution to the 450 μl elution and incubate overnite in a 65C heat block

C.3.4 Final DNA cleanup

Takes a morning. Make sure to let the ethanol precipitation sit for a long time and spin for a LONG time to pellet/precipitate the short DNA fragments. I proceed similar to the section **Quantification Part 2** above, except I add an extra 10 min to each centrifuge time, just to be safe.

1. add 10 μl EDTA, 20 μl Tris [pH 8] then add 1 μl of proteinase K and incubate 1hr at 45C.
2. separate DNA from proteins by phenol:chloroform extraction using gel phase lock tubes

3. add 1/10th volume NaAcetate and 2 μl of N mg/ml glycogen as a DNA carrier and ethanol precipitate

4. resuspend DNA in 100 μl of TE

Quantification of the product from the above protocol doesn't seem to be useful and just wastes a little sample. There's not enough DNA to quantify.

C.3.5 qPCR to determine TF binding site enrichment

Even with everything in plates, setting up the entire 384-well plate (i.e. using most or all of the wells) takes 1hr 30min.

qPCRs are run with an ABI Sybr Green master mix in a 384-well plate using an ABI 7900HT qPCR machine. The qPCR is a 2-cycle PCR with 60C melting temp so primers should be designed accordingly (e.g. with Primer3 software).

Use 1.5 μl of template (from the total 100 μl), 150 nM primer, 10 μl 2x master mix and add water to a final volume of 20 μl . I dilute the primer and the template quite a bit in water to make it easier to pipette with the multichannel (I use 1.5μl as the smallest pipetting amount). Primers are typically prearranged in 96-well plates so everything (primers, template, master mix) is added 12 wells at a time with a 12-channel pipettor (Finnpipette). Make sure to have more than enough of everything or else it is hard to get consistent volumes from the multichannel. Also making sure *every* tip is well attached to the pipettor is important for consistency reasons, otherwise they show up half-full.

C.4 ChIP Protocol Post 1st Round Factorial Optimization

I took the ChIP protocol above and optimized it through one fractional factorial experiment with eight factors (see section 2.2 on page 110 for details). This isn't the final protocol, but it will probably be used to verify some predictions from Vwani's lab at UCLA, because this protocol is much faster and less taxing than the original protocol above that I used for the PLoS paper.

C.4.1 Lyse, Crosslink, Shear

The following protocol should be started in the morning and will finish later on in the evening (allocate 10hrs). The morning will be only sporatic work but as the day goes on and on, full time must be devoted to the experiment (from the lysis step forward). Little has changed in this part of the protocol except the additional glycine quenching step.

Growth

1. Grow stain(s) in 50ml of liquid media in a 250ml flask with appropriate antibiotics

2. add 0.01 μM IPTG or other inducer as needed for induction of any cloned constructs (if you have a specific non-cloned gene or tagged gene in the genome, this step is not necessary)

3. put a 15 ml sample from the flask into a 15ml corning centrifuge tube.

Crosslinking

1. add 37% formaldehyde to 1% total concentration for each (400μl for 15ml sample) and mix by inversion (invert tube 10 times). Incubate 10 min at room temp. Do NOT crosslink too long!
2. quench the crosslinker with 1/20th volume of 2.5M glycine (750 μl for a 15 ml sample)
3. pellet cells by centrifugation for 10 min at 3200g (this doesn't pellet ALL of them but enough for downstream purposes; if OD is low (< 0.3) all of sample will pellet)
4. decant supernant in formaldehyde/media hazardous waste bin
5. wash cells 2x in ice-cold PBS (spin 3200g for 8 min) (from here on, keep samples on ice to prevent proteases from messing up the experiment; I use 2/3 initial volume for washing: e.g. with 15ml of cells I use 10ml of PBS to wash)
6. prior to lysis remove any remaining drops of PBS that wouldn't go away by decanting with a P1000 pipettor

Lysis and Shearing Do these the same as in the previous protocol.

Quantification Part 1

1. remove 100 μl from each sample to quantify the amount of DNA in each lysate and to verify the shearing range
2. place the remaining 900 μl in the -80C freezer for use in immunoprecipitation (it'll be more than enough)
3. add 25 μl of water to the 100 μl sample (125 μl total volume); add 5 μl of 20mg/ml proteinase K;
4. reverse crosslinks overnite in a 65C heat block or water bath

C.4.2 Sheared DNA yield, immunoprecipation, and bead washing

This section is the largest change to the previous protocol. These steps used to take at least 2 days, now you can easily finish them in one - if you work hard you can even reverse the crosslinks and have the DNA cleaned up at the end of this day.

Quantification Part 2

I found that a Qiagen PCR purification actually seems to be a more consistent way of determining the relative yields of your sheared DNA. Depending on how robust the ChIP proceedure is to the concentration of DNA, I might eliminate this step altogether, and use it only as a sanity check on the shearing range.

1. clean up sheared DNA with a Qiagen DNA purification kit
2. resuspend DNA in 30 μl of EB buffer
3. quantify using spectrophotometer (e.g. a Nanodrop)

500-600ng of each sheared DNA sample should be run an a 1.5% agarose gel to verify shearing range. Examples of yields vs OD can be seen on pages 95 and 102. For the next step it is helpful to have the DNA concentration over 250 ng/μl . I no longer use this DNA as a positive control in my qPCR rxns, because I never used that data and it was just an unnecessary cost.

Immunoprecipitation and bead washing

This step is much faster in the new version. Rather than doing an overnite antibody incubation + 2 hrs bead incubation, I do a 10 min antibody and a 10 min bead incubation. I also removed the preclear step, because it had little to no effect (if anything it was deleterious). I switched from agarose beads to dynal beads, because the performance was similar, and the dynal beads require less brain power and attention to wash.

Immunoprecipitations are begun with equal starting DNA (25 μg). So DNA is first quantified using the protocol above. The following steps I do in 1.5 ml eppy tubes (I used to use a 2.0 ml eppy tube, but the 1.5 works better with the dynal beads).

1. divide the 900 μl left of the sheared lysate into N replicates where N is divisible by 2. Each sample should contain 25 μg of DNA. Label these **B**, and **C** (I normally only use N=1).
2. dilute all samples 1:10 in dilution buffer (if your DNA isn't concentrated, dilute as much as possible, should still work).
3. add 2 ug of the correct antibody to sample **B** (e.g. if your tag is myc put anti-myc here)
4. I no longer use the incorrect antibody negative control. sample **C** is just a no antibody negative control
5. rotate all samples 10 minutes at 4C (in the cold room), preferably in 1.5 ml eppy tubes

By pulling both of your immunoprecipitations from the same lysate you get nice samples for estimating the enrichment between B, and C. Independent sample replicates are important as well. Just try to start with the same concentration of DNA in every immunoprecipitation.

Wash beads

With the switch to dynal beads, it is much more difficult to remove the beads by mistake. I do each wash with 1 ml. And I remove the washes with a P1000 pipettor. I do all of this work in the cold room to save time whilst still keeping the samples cold.

Rotate each wash 5 min in cold room, except for TE washes, which are rotated at room temp.

I prepare the dynal beads using a similar strategy to the Young Lab protocol in Nature Protocols.

1. add N μl of beads
2. collect with magnet;
3. add 15 x N μl of block solution (0.5% BSA in PBS)
4. repeat steps 2 and 3;
5. resuspend beads in N μl of block solution

This can be done during the first antibody incubation. It doesn't hurt if that incubation goes too long (I saw no difference between 2 hrs and 10 minutes).

Add beads

add 60μl of the prepared dynal protein G magnetic beads [invitrogen]. Rotate at 4C for 10 minutes.

Bead Washing

1. wash beads 1x with low salt wash
2. wash beads 1x with high salt wash
3. wash beads 1x with LiCl wash
4. wash beads 2x with TE (steps at room temp from now on)
5. For the elution, I used the Young protocol (which used something similar to TE + 1% SDS) for the dynal elution. Crosslinks were reversed overnight in a water bath at 65C (but the crosslinks are reversed after 6hr if you want to continue in the same day)

Crosslink reversal

1. add 1 μl of proteinase K samples the next morning after taking them out of the water bath To the dynal samples I added H_2O to 450 μl total volume for phenol chloroform extraction

DNA cleanup and qPCR This steps are still done the old way with phenol/chloroform and 20 μl qPCR rxns. Instead of having four samples (+ control, correct antibody, 2 x negative controls) per tested edge, I now only use two, which cuts the expensive qPCR rxns in half and still yields the same results.

C.5 Preparation of *E. coli* genomic DNA

The following protocol is modified from this website, which modified the protocol in *Experimental Techniques in Bacterial Genetics, Jones and Bartlet 1990*. It will produce purified DNA suitable for PCR of genes, promoters, etc... (much better than just tossing in some *E. coli* cells into your PCR), but the DNA will be fairly sheared. If you want contiguous chromosomal DNA it is better to use other protocols.

Gotchas: Genomic DNA is a pain to work with. It is really thick, sticky, fragile, hard to pipette and very hard to resuspend. If I don't need long contiguous pieces (which is almost always the case), I shear it by sonication to make it easier to handle. If you don't shear, YOU MUST WAIT A LONG TIME for the DNA to go back into solution after you precipitate it. Wait at least 24 hr for the DNA to go into solution and mix well otherwise it is easy to have a non-uniform distribution of DNA in your solution.

1. Grow 5ml *E. coli* overnight in rich broth.

2. Transfer 2ml to a 2ml eppy tube (or a 1.5 if that's the biggest you have)
3. pellet cells by centrifugation for 60 sec at 5600g
4. vacuum off supernatant
5. resuspend in 482 μl TE[2]. Add 15 μl of 20% SDS and 3 μl 20 mg/ml proteinase K and incubate 1 hr at 37C.
6. *optional*: sonicate the DNA to make it easier to handle a Branson 250 Sonifier for 30 secs at 10% power (higher power than this in the 2 ml tube results in excessive foaming and not very much shearing e.g. Figure 5.5, page 214)
7. add 500 μl phenol/chloroform and mix well
8. prepare a phase-lock gel (light) tube by spinning at max speed for 30 sec
9. transfer the mixed solution to the prepared phase-gel lock tube
10. spin at max speed for 10 min
11. transfer the aqueous layer (the part above the gel) to a fresh 1.5 ml tube
12. add 5 μl RNAse Cocktail
13. incubate at 37°C for 25 min
14. add 500 μl phenol/chloroform and mix well
15. prepare a phase-lock gel (light) tube by spinning at max speed for 30 sec
16. transfer the mixed solution to the prepared phase-gel lock tube
17. spin at max speed for 10 min
18. transfer the aqueous layer (the part above the gel) to a fresh 1.5 ml tube
19. add 50 μl sodium acetate
20. add 500 μl isopropanol and mix gently (you should see the genomic DNA in 1-2 minutes if you didn't shear it, wait 3-5 minutes total before preceeding)
21. spin at max speed for 4 minutes
22. remove supernatant with a weak vacuum (I use -200 mbar)
23. add 1 ml of 70% ethanol and incubate at RT 1 minute
24. spin at max speed for 2 minutes (you want the DNA pellet to stick to the tube, unfortunately with genomic DNA it often is more of a gooey ball that won't stick so be careful. If I can get the DNA to stick, I suck as much liquid as I can without getting close to the pellet and add more ethanol (filling the tube) and mix well. This seems to remove some of the gooeyness and allow the pellet to stick.
25. remove supernatant with a weak vacuum (I use -200 mbar)
26. resuspend in 200-500 μl TE

[2] if you don't want *really* sheared DNA try to be gentle with your resuspension (i.e. don't vortex the crap out of the cells)

C.6 Preparation of *E. coli* plasmid DNA

C.6.1 Standard Method

I almost always use the Qiagen Miniprep kit. I've also used the Eppendorf Miniprep kit but don't like it as much.

C.6.2 Old School Method

This method is very similar to the Qiagen procedure except you have to make all the solutions for yourself. It is quite a bit slower. Yield???

1. Grow 5ml *E. coli* overnight in rich broth.
2. Transfer 2ml to a 2ml eppy tube (or a 1.5 if that's the biggest you have)
3. pellet cells by centrifugation for 30 sec at 7500g
4. vacuum off supernant
5. repeat above two steps with an additional 2ml of cells in the same tube (only necessary if you want a lot of DNA)
6. resuspend in 100 μl of GTE buffer (50 mM Glucose, 25 mM Tris-Cl, 10 mM EDTA, ph8). Vortex gently if necessary.
7. Add 200 μl of NaOH/SDS lysis solution (0.2 M NaOH, 1% SDS). Invert tube 6-8 times (solution should become very quickly)
8. Immediately add 150 μl of 5 M potassium acetate solution (pH 4.8) to neutralze the NaOH from the previous step and precipate the genomic DNA and SDS into a white goopy mass. Spin at max spin for 1 minute.
9. Transfer supernant to a new tube; don't transfer any of the white junk.
10. precipitate with 500 μl isopropanol on ice for 10 minutes and centrifuge at 4C for 2 minutes
11. aspirate *all* of the isopropanol supernant. Dissolve the pellet in 400 μl TE.
12. phenol/chloroform
13. add 50 μl NaAcetate
14. ethanol precipitate
15. resuspend in 50 μl TE

C.6.3 RNA-free Midiprep

I found this protocol on the internet, and I've added my own personal comments about the tricky steps.

1. Grow 60 ml of cells in rich media (e.g. LB) with appropriate antibiotics to maintain the plasmid. Grow them to late stationary (high cell density).
2. transfer 50 ml of cells to a 50 ml falcon tube
3. spin at 4000 rpm[3] for 15 min at 4°C [4].
4. Add 20 ml of H2O and mix a little (I don't resuspend the entire pellet; not a good wash, but it removes most of the LB which can mess things up). Spin 4 min at 4000 rpm[5]. Remove all of the solution (suck out most of if with a vacuum and get the last little bit with a pipettor).
5. Resuspend the pellet in 5 ml of Solution I (see B.5.1)
6. Add 10 ml of Solution II (see B.5.2) and mix by inverting the tube 10 times (solution should become clear)
7. Add 7.5 ml of Solution III (see B.5.3) and mix by inverting the tube 10 times (solution should fill with chunky white stuff)
8. Spin at 4000 rpm for 15 minutes[6]. Since the centrifuge I use doesn't reach the ideal speed, not all of the white stuff (genomic DNA and cell wall) pellets to the bottom. However, most of the precipitate does go to the bottom and the remaining part sits as a firm layer on the top of the tube and the following step is still pretty easy to accomplish.
9. Transfer the supernatant to a new 50 ml falcon tube (avoid taking the white stuff).
10. Add 15 ml of isopropanol, mix well, and store at room temperature for 10 minutes.
11. Spin at 4000 rpm for 15 minutes at 4 C. Discard the supernatant. Remove any remaining fluid with a pipettor.
12. Dissolve the pellet in 600 μl TE. Transfer to a 1.5 ml eppy tube.
13. add 200 μl of 8M LiCL. Mix well and spin at 14,000 rpm for 5 min at 4°C (this precipitates the larger RNAs so you can get rid of them).
14. transfer the supernatant containing the plasmid DNA to a new 1.5 ml eppy tube. Add 600 μl isopropanol. Mix well and incubate 2 minutes at room temperature. Spin at 14,000 rpm for 5 minutes at 4°C.
15. discard the supernatant. rinse the pellet and the wall of the tube (by inverting it a few times) with 1 ml of cold 70% ethanol.

[3] this is the fastest our centrifuge will go, the actually protocol says 6000 rpm for 10 min

[4] not all, but most of the cells will be pelleted; solution is still cloudy, so longer spin times might bring up the final DNA yield

[5] the original protocol reads: (optional) Resuspend the cells in 20 ml of H2O. Spin again.

[6] original protocol recommends 10 min at 10,000 rpm

16. add 400 μl of TE with RNase A (20 ug/ml) (I actually use Ambion RNAse cocktail; just add 2μl of the cocktail). Incubate 30 minutes at 37°C (this chops up the remaining RNA into little bits)

17. After 30 min, if a nucleic acid pellet is visible at the bottom of the tube, vortex well to dissolve and incubate another 30 minutes

18. add 240 μl 2M NaCl, 20% PEG8000 (10 g PEG 8000 and 5.844 g NaCl in 50 ml H2O) (I think this part removes small bits of RNA, but I'm not totally sure)

19. spin at 14,000 rpm for 5 minutes

Gotchas: The step below might make you nervous because the pellet becomes invisible. There is probably a ton of DNA there, but don't worry that you can't see it. If you can't see it after the addition of ethanol afterwards, then you can worry (because the ethanol precipitated DNA should be white as it normally is).

20. Discard the supernatant. Rinse the pellet with 300 μl of cold 70% ethanol (pellet will become white but is much smaller than before due to the absense of the RNA).

21. resuspend the pellet in 400 μl of TE

22. add 400 μl of phenol:chloroform and mix well

23. spin 5 minutes at max rpm and keep the aqueous phase

24. add 400 μl of chloroform, mix

25. tranfer the solution to a gel phase lock (Light) tube. spin 5 min max rpm, decant the aqueuous phase to a 1.5 ml eppy tube

26. add 40 μl 3 M sodium acetate and 880 μl of 95% ethanol. Mix well. Spin at max rpm for 5 minutes at 4°C.

27. Discard supernatant. Wash with 1 ml 70% ethanol.

28. Discard supernatant, allow ethanol to evaporate

29. resuspend in 200-500 μl TE (I usually use 250 μl)

C.6.4 Midiprep

Gotchas: Not really a gotcha, but I've given up on this protocol which yields RNA rich, dirty samples. The RNA-free Midiprep protocol above is much better.

For 20-50 ml, resulting in quite a lot of plasmid DNA. Grow cells in LB with appropriate antibiotics. I typically grow 50 ml in a 250 ml flask. This protocol is from the Sambrook molecular cloning manual.

1. add 15 ml of overnite to a 15 ml centrifuge tube
2. spin at max speed (4000 rpm) in the bucket centrifuge for 10 min at 4 C
3. aspirate the media (according to Sambrook it is very important to get the pellet very dry to prevent DNA that is hard to cleave with restriction enzymes)
4. resuspend pellet 200 μl Alkaline Lysis I (see B.5.1) by vortexing
5. tranfer the 200 μl to a 1.5 ml eppy tube
6. add 400 μl of Alkaline Lysis Solution II (see B.5.2 *make fresh each time*)
7. invert tube rapidly 5 times [7]
8. *immediately* add 300 μl of ice-cold Alkaline Lysis Solution III (see B.5.3)
9. centrifuge at max speed for 5 min
10. transfer 600 μl of the supernatant to a fresh tube
11. add 5 μl of RNAse cocktail and incubate at 37°C for 25 min
12. add an equal volume of phenol:chloroform to the tube and mix by vortexing.
13. centrifuge at max speed for 2 min
14. add 600 μl isopropanol at RT and 2 min
15. centrifuge at max speed for 5 min
16. remove supernatant
17. add 1 ml 70 % ethanol
18. remove supernatant and dry
19. resuspend in 100 μl TE

C.7 Preparation of *E. coli* RNA

C.7.1 RNAeasy preps

How much sample to use

This is what I do with *E. coli*.

OD	culture (ml)	RNA protect (ml)	RNAlater (ml)
0.5	2	4	
0.25	4	8	

[7] make sure the entire tube surface is coated with Soln II

C.8 Size-separation / exclusion of DNA

C.8.1 Size-exclusion of DNA using microcon filters

Microcon filters are ultrafiltration columns that can be used to remove salts, concentrate DNA, and remove DNA less than a particular size (125 bp maximum double-stranded and 300 bp maximum single-stranded). This isn't as complete a removal as gel filtration. Each of the nucleotide cutoffs below indicates which lengths retain at least 90% of their molecular species. So it is almost always better to go with bigger sizes if the pieces you want to retain are much larger than the smallest cutoff.

The following table is copied from the millipore website: http://www.millipore.com/publications.nsf/docs/6dkp6d

NMWL	single-strand cutoff (bp)	double-strand cutoff (bp)
3K	10	10
10K	30	20
30K	60	50
50K	125	100
100K	300	125

Here are their recommended g-force and spin times for the microcon columns:

NMWL	max g-force	spin time (min) at 4°C	spin time (min) at 25°C
3K	14,000	185	95
10K	14,000	50	35
30K	14,000	15	8
50K	14,000	10	6
100K	500	25	15

Note that these columns also retain proteins which are larger than the pores in the filter, so this isn't a good way to remove proteins from your reaction.

C.8.2 Size-exclusion using Qiagen spin-columns

The Qiagen columns use the glass fiber to catch the DNA and wash the salts and proteins off. The sample loss is much higher than with an EtOH or a microcon filter. But it denatures proteins and should remove them better.

From the Qiagen website:

Specifications:	PCR purification kit	nucleotide removal kit	gel extraction kit
Recovery:			
Oligonucleotides	-	17-40mers	-
dsDNA	100bp - 10kb	40bp - 10kb	70bp - 10kb
Removal:			
<mers	Y	Y	Y
17-40mers	Y	N	N

C.8.3 Size-separation of DNA using Sephacryl 500

I've taken a

C.9 Preparation of PET libraries

C.9.1 Growing cells

You want to grow your cells to almost max out the RNAeasy column in the RNA step below (max is 100 μg). If you don't, you will have problems getting enough cDNA downstream, particularly because the rRNA removal step removes most of the RNA.

For *E. coli*, I grow the cells to around 0.5 (OD600) and add 2.5 ml of this to 5 ml of RNAprotect, vortex 5 sec, incubate at RT for 5 minutes, and spin at 4000 rpm in a bucket centrifuge for 12 minutes. Then I decant the RNAprotect and get the residual off by tapping the tube on a paper towel. RNAprotected RNA is safe in the -20C for two weeks (I try to use it ASAP though).

C.9.2 RNA extraction

When making PET libraries, I really want to remove every trace of DNA possible from the RNA before making cDNA. Otherwise, if I unknowingly have DNA in there it would create lots of false positives when defining genes with the sequenced PETs. Therefore, I use the RNAeasy kit (preferentially selects RNA over DNA, but still has a lot of DNA left over; also removes short RNAs). I then precipitate the RNA with LiCl, which does not precipitate DNA (this removes the bulk of the remaining DNA and short RNAs). Finally, I digest the trace remaining DNA with the Ambion DNA-free kit.

The entire RNA extraction process takes 4-5 hours (?)

1. Lyse cells in 100 μl of TE with 1 mg/ml lysozyme. Incubate 2 min, vortex every minute. Add 10 μl Proteinase K. Incubate 3 more minutes, vortex every minute.
2. add 350 μl RLT (with β-ME added) and follow the RNAeasy kit; elute with 50 μl 2 times (100 μl total)
3. SAMPLE POINT A: measure yield with Nanodrop, save 750 ng for a gel (takes approximately 1 hr to reach SAMPLE POINT A)
4. add 50 μl (1/2 volume) of 7.5 LiCl [ambion] to the 100 μl of RNA; place at -20C for 30 minutes
5. centrifuge at max rpm for 15 minutes
6. wash RNA pellet in 1 ml of 70% ethanol; incubate at RT 2 minutes, spin 5 minutes, dry pellet 7 minutes
7. resuspend in 35 μl of RNAse free TE [Ambion] [8] (Note: It takes 1 hr to get to this point from SAMPLE POINT A)
8. follow DNA-free TURBO kit instructions for high-conc DNA. Briefly: add 3.5 μl Buffer, add 1 μl DNAse, incubate 30 min, add additional 1 μl DNAse, incubate 30 more minutes. Deactivate with 7 μl of deactivation buffer and keep supernatant.

[8] In the past, I resuspended into 50 μl , which is better for the DNA-free kit. However, the MICROBExpress kit allows at most 10 μg in 15 μl . With 50 μl it wouldn't have been concentrated enough to get 10 μg in such a small volume. I want to make sure and maximize the starting material, because so much RNA is lost after the rRNA removal (final yield from 10 μg is expected to be 1-2.5 μg).

9. transfer the upper, aqueous phase to a new eppy tube
10. SAMPLE POINT B: spec DNA free RNA and save 750 ng for a gel (it takes 2 hr 30 minutes to get here from SAMPLE POINT A)
11. use MICROBExpress to remove 16S and 23S from 10 μg of total RNA (max volume 15 μl).
12. resuspend in 16 μl (not the recommended 25 μl)
13. SAMPLE POINT C: spec
14. save 200 ng to run on gel (more if possible?)

C.9.3 1st strand synthesis of cDNA

This step takes about 2 hrs

Use Superscript III and the following protocol:

Do in PCR tubes:

1. add 1 μl of random hexamers (100 ng)
2. add 1 μl of dNTP (10 mM each)
3. add 1.5-3μg of *mRNA* [9]
4. add H_2O to 13 μl
5. heat to 65°C for 5 minutes, chill on ice, brief centrifuge
6. add 4 μl First-strand buffer, 1 μl DTT
7. add 1-3 μl of SuperScript II, mix by flicking tube a few times [10]
8. incubate at 25°C for 5 minutes to bind random primers
9. incubate at 50°C for 60 minutes
10. heat-inactivate at 70°C for 15 min

C.9.4 2nd strand synthesis of cDNA

This step take 3 hours

Do this in the same tube as first strand. Keep on ice while preparing.

1. add 66.15 μl of H_2O
2. add 10 μl of NEBuffer 2

[9] this is more than Invitrogen recommends for mRNA, but this is important to have enough cDNA for downstream steps. Note that 11 μl is as much RNA as you can add, so it helps to have it really concentrated.

[10] add 200 U (1 μl) of Superscript III per 1 μg of mRNA

3. add 3 μl dNTP mix (10 mM each)
4. add 5 μl *E. coli*DNA polymerase I (40 Units)
5. add 0.25 μl RNAse H (1 Unit)
6. incubate 2 hours at 16 C
7. add 5 μl E. coli DNA ligase buffer (NOT T4 ligase buffer)
8. add 1 μl E. coli DNA ligase (NOT T4 ligase) and add another 0.25 μl of RNAse H (1 Unit)
9. incubate 15 minutes at 16 C
10. heat inactivate both enzymes 20 min at 75 C (it takes about 11 hrs from the beginning to reach here)
11. (I go home for the day after starting the previous step. I set the thermocycler to keep the tubes at 4C until the next morning)
12. cleaned up with Qiagen PCR clean up; eluted into 35 μl EB buffer [11]
13. end repair with epicenter kit using 34 μl cDNA (all of it; just keep the same tube); incubated at RT 45 min
14. heat deactivated enzymes 70 C for 10 min
15. clean up the end-repaired DNA with a phenol:chloroform cleanup; and elute into 30 μl TE buffer
16. SAMPLE POINT D: spec 1 μl

C.9.5 Add adaptors to double-stranded cDNA

Use 2 μl of N mM adaptor pairs in each reaction (appx 4.2 μg). Anneal them first in TE+salt.

1. to the 29 μl of cleaned up, end-repaired DNA (1 μl was used to spec), add 3.6 μl T4 DNA ligase buffer
2. add 2 μl (appx 4.2 μg) of BamISH adaptor (see section 6.6.1 page 267 for details on BamISH)
3. add 1 μl of T4 DNA ligase
4. mix by flicking the tube a few times
5. incubate for 12 hrs at 16°C
6. heat inactivate T4 ligase at 65 C for 10 min
7. add 1 μl of T4 DNA ligase buffer[12]

[11] 35 μl was chosen because it allows 1 μl to be used to spec the DNA and the remaining amount is the maximum allowable volume for the end-repair kit

[12] in case the ATP has been exhausted from the long ligation

8. add 1 μl of T4 polynucleotide kinase (no need to add ATP because it is in the ligase buffer) [13]
9. incubate at 37°C for 30 minutes
10. heat inactivate for 20 minutes at 65°C
11. clean up with Qiagen PCR purification kit, elute into 30 μl

clean up now or just run on gel?

C.9.6 Size-selection of cDNA

We need to remove the primers and to only grab the longer cDNA as the short pieces are preferentially amplified by RCA. We'd prefer to have longer cDNA as they should be closer to full length genes/operons. I select two sizes from the gel 500-1500bp and $>$ 1500 bp.

[13] we are adding phosphates to the adaptors

Appendix D

Equipment

D.1 DNA work

Figure D.1: We have two BioRad PCR machine both with two 48-well blocks. If life were to run my way, we'd have one with a 96-well block and one with a 384-well block.

Brief Update *Sat Jul 29 20:14:37 EDT 2006*: Life is now running my way with regards to PCR machines. We now have a 384-well and a 96-well block. Gear that works in them is (all from BioRad): MLP-9601 unskirted PCR plate (for 96-well PCR min vol 10 μl), MSB-3842 (for 384-well min vol 5 μl), MSB-1001 microseal 'B' adhesive seals (for the top of either plate).

D.1.1 qPCR

The following instructions are for the 7900HT. The template was setup by David Lorenz in Jim Collin's lab.

Josh Thaden's 12 steps to qPCR success

1. spin plate (1000 RPM for >= 2 min) **IN A ROOM TEMPERATURE CENTRIFUGE**[1]
2. log into computer log: collins, pwd: genetream
3. open *PCR-w-melt* icon
4. define wells and click *Use*
5. set sample volume (instrument tab, to the right of the define wells tab)
6. save data in appropriate folder
7. click *Connect* (instrument tab in the main menu) [2]
8. Click open/close button in Tools menu; Load 384-well plate
9. Click open/close button (to close door of machine)
10. press *Start* in instrument tab
11. software inferace will show remaining time after a couple minutes
12. Save two files (go to *Export*): results table (just Ct values), clipped (fluorescent values at each time)

D.2 Gels

D.2.1 MiniGels

Protein

Run using premade gels from invitrogen using their sure-lock system.

Nucleic Acid

Run using fisher mini gel rigs

[1] if the centrifuge is not at ambient temperature, you'll get condensation on all of your wells; for this reason, most users prop open the lid on the centrifuge in the qPCR room

[2] if you are using the hs dynamo kit from NEB, you need to switch to 15 minutes initial heat denaturation

D.2.2 Big Gels

We have an Owl Gater A1 gel rig (the owl rigs are very nice).

gel thickness (cm)	0.25	0.5	0.75	1.0
volume to achieve thickness above (ml)	81	163	244	325
1.0 mm, 25 well comb volume (μl)	2	7	12	17
1.5 mm, 25 well comb volume (μl)	3	11	18	26
1.5 mm, 9 well comb volume (μl)	8	28	49	69

Agarose gels are run

D.2.3 Polyacrylamide pre-cast Gels

I use the ones from invitrogen:

well/comb type	**maximum volume**
5 well, 1.0 mm	60 μl
8 well, 1.0 mm	28 μl
10 well, 1.0 mm	25 μl
10 well, 1.5 mm	37 μl
12 well, 1.0 mm	20 μl
15 well, 1.0 mm	15 μl
15 well, 1.5 mm	25 μl

For TBE Gels I use Hi-Density Sample Buffer (5x) for 6% gels the bromophenol blue (dark blue) dye band is at 65 bp and the xylene cyanol band (blue green) is at 250 bp. I run the 6% TBE gel at 200V for (30-90) minutes.

For TBE Urea gels I use Novex TBE Urea Sample buffer and I flush the wells a few times with 1x TBE running buffer to remove Urea. I also heat the samples for 3 minutes at 70°C after mixing the sample with the loading dye. The recommend not more than 200 ng DNA / band. Use standard TBE running buffer.

I use the elution buffer mentioned in Science Shendure *et.al.* polony paper to remove the DNA from cut bands (this is pretty similar to the sambrook method, *Isolation of DNA Fragments from Polyacrylamide Gels by the Crush and Soak Method* protocol 12, 5.51): dice fragments (I use the tube inside a tube method, NOT the razor blade technique they mention). Then I add PAGE elution buffer to the diced pieces: 10 mM Tris, 50 mM NaCl, 1 mM EDTA. Place at 37 C overnight. Next day spin down 1 min maximum speed. To improve yield wash gel framents with an additional 200 μl of elution buffer. Purify with phenol:chloroform.

Below is taken from sambrook and the invitrogen novex quick reference card (a * indicates invitrogen sells that size and that I took the dye migration numbers form invitrogen) the number in parenthesis indicates the TBE-Urea dye migration band (see also Figure D.2):

concentration of acrylamide	effective range of separation (bp)	bromophenol blue (dark)	xylene cyanol (light)
3.5%	1000-2000	100	460
5%	80-500	65	260
* 6%		65 (25)	250 (110)
* 8%	60-400	25	220
* 10%		35 (20)	120 (55)
12%	40-200	20	70
* 15% (urea only)	25-150	15 (10)	60 (40)
* 20%	6-100	15	50

D.3 Imaging

Gels are imaged on a Versadoc imaging system.

D.4 Weighing, Measuring, pH

D.5 Growing stuff and warming stuff

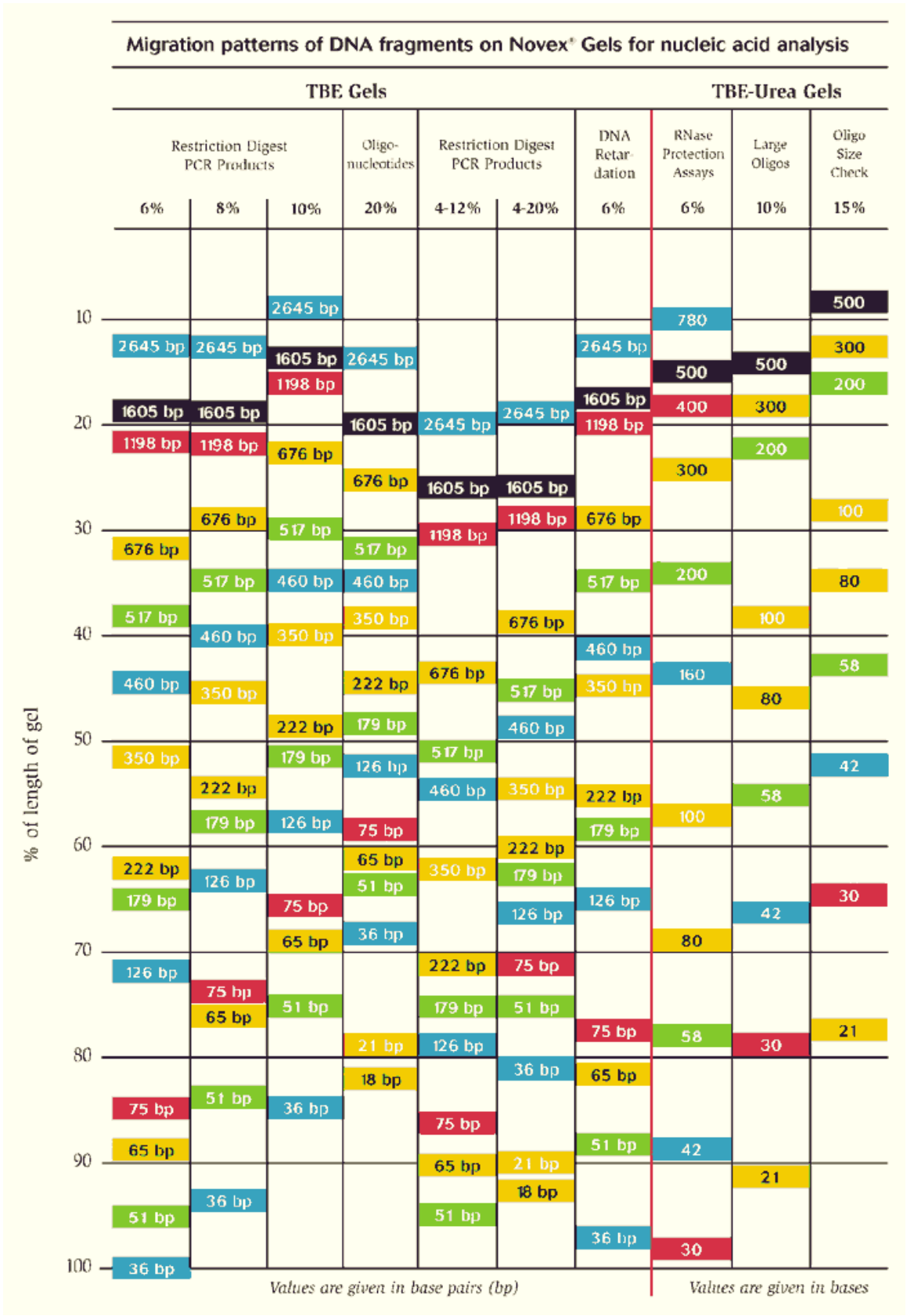
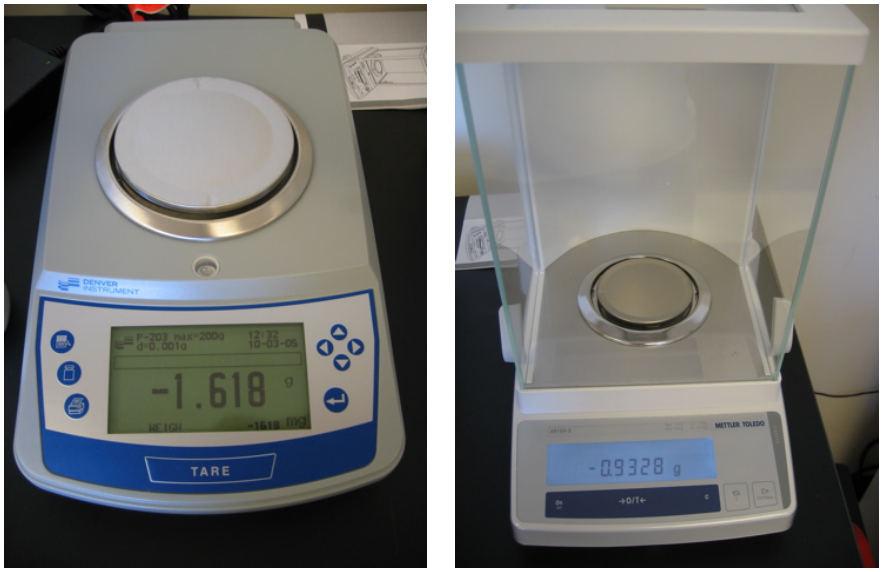


Figure D.2: Gel migration chart for the novex gels.



(a) Denver Instruments scale (b) Toledo AB104-S scale

Figure D.3: We have two scales. The Denver (a) is for low precision stuff and the Toledo is for higher precision weighing. I use the Toledo for anything under a gram. If I were to buy a low precision scale. I wouldn't buy this Denver, it maxes out at 200g. If you have a deep-well plate that you are having trouble balancing on the centrifuge it can easily weigh more than this. I get a scale that goes up to 5kg even if I had to give up a little precision

Figure D.4: The Denver 200 pH meter does the trick

Figure D.5: This Thermo electron shaking incubator is very quiet but only works well at temperatures 35C and above. It is also *very slow* to start and stop and has an annoyingly long delay between when you push the start/stop button and when it decides "ok I'll begin stopping now". We have two stackable ones.